中国运筹学会
中国工业与应用数学学会 丛书

应用数学译丛

第2号

概率论和随机过程

〔日〕伏见正则 著
李明哲 译

世界图书出版公司

概率论和随机过程

[日] 伏见正则 著

李明哲 译

世界图书出版公司

北京 · 广州 · 上海 · 西安

1 9 9 7

图书在版编目(CIP)数据

概率论和随机过程/(日)伏见正则著;李明哲译. —
北京:世界图书出版公司北京公司, 1997.2
(应用数学译丛/章祥荪等主编)
ISBN 7-5062-2852-1

Ⅰ.概… Ⅱ.①伏…②李… Ⅲ.①概率论②随机过程
Ⅳ.0211

中国版本图书馆CIP数据核字(96)第23379号

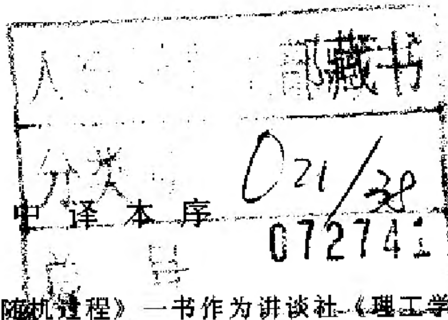
伏见正则
确率と确率过程
株式会社講談社,1987

概率论和随机过程

[日] 伏见正则 著
李明哲 译

责任编辑 高 蓉
世界图书出版公司北京公司出版
北京朝阳门内大街137号
北京市昌平县百善印刷厂印刷
世界图书出版公司北京公司发行 各地新华书店或外文书店经售
*
1997年2月第 1 版 开本:850×1168 1/32
1997年2月第1次印刷 印张:6.5
印数:0001-2500 字数:17万字
ISBN: 7-5062-2852-1/O· 174
著作权合同登记 图字:01-95-503号
定价:15.00元





中译本序

日语版《概率论和随机过程》一书作为讲谈社《理工学者所著数学丛书》分册于 1987 年首次面世，出版这套丛书的目的如下：

传统认为，高等数学理论应由数学专家讲授，但数学作为物理及工程技术领域的应用工具，有时并无此必要. 这套丛书以直观阐述为主，力求从工学的角度描述数学理论. 它共分七册，包括线性代数，常微分方程，偏微分方程，多变量及复变函数的微积分学，傅里叶解析 (Fourier analysis)，概率论等内容. 它的定义、定理和内容构成都着眼于读者，为简洁明了地解释纯数学理论做了很大努力.

这套丛书在上述出版方针下，得到了广大以应用为目的而学习数学的学生们和研究人员们的支持，多次再版发行.

这次，借中国科学院应用数学所所长章祥荪教授访问我所属的东京大学工学部计数工学科之际，由他介绍，使得我的书能够由世界图书出版公司北京公司出中文版，感到非常高兴. 对此，我向章教授和该出版公司表示深深地谢意. 正在我的研究室攻读博士学位的李明哲同学花费了大量时间和精力致力于本书的翻译工作和繁琐的计算机输入工作. 没有他的献身精神和努力，也不可能完成这次出版. 对此，我也向他表示深深地谢意.

我于 1991 年 8 月在北京召开亚洲及太平洋地区运筹学国际学术会议 (APORS' 91) 之际，第一次访问中国. 时隔四年，再借运筹学及其应用国际学术研讨会 (ISORA'95) 之机访问北京，并看到本书中文版的出版工作正在顺利进行之中，感到由衷高兴. 希

望本书能够对中国的广大读者起到一定的帮助作用.

伏见正则

1995年8月于东京

译 者 序

对于工程技术人员来讲，数学知识是手段，实际运用是目的．目前国内外流行的许多应用数学参考书，多半偏重于数学方法的论证，而对于如何让这些读者在一定时间内，正确广泛地了解和掌握数学知识不够重视．

在这种背景下，这本书应运而生，并在众多的应用数学参考书当中得到了瞩目，获得了许多学者们的高度评价和大量读者的喜爱欢迎．东京大学名誉教授、著名学者森口繁一先生曾为作者能够在有限的篇幅内生动准确地叙述如此丰富的内容而感叹，京都大学名誉教授、著名学者伊藤清先生也为作者深厚的数学功底和多年的教学经验所折服．我本人在学习这本书的过程中对此深有体会，并决心把这本书介绍给广大的中国读者．

本书的出版，是在中国科学院应用数学研究所所长章祥荪教授的亲切关怀和帮助下得以完成的．在此对他表示衷心的感谢．在本书的翻译出版过程中，中国科学院数学研究所的许鹏程同学和北京大学计算机研究所的邓集锋同学、东京大学大学院数理科学研究科的梁淞同学做了许多整理工作．最后，中国科学院数学研究所的井竹君教授对译稿作了仔细的校阅．在此对他们也一并表示诚挚的谢意．

李明哲

1995 年 10 月

序

大约在十七世纪，概率论出于对赌搏的关心而作为解释随机现象中所存规律性的方法开始得到研究，迄今已被抽象成一套严谨的数学理论. 特别是最近，研究的主流已开始转向解析因时间而发生变化的随机现象，即随机过程论.

现存的大多数概率论和随机过程教本，对以应用为目的的初学者来讲，可能会感到有些吃力. 这本书轻松易懂，可做为学习纯理论教本之前的入门书使用.

基于这一点，本书把重点放在直观地理解而不是严谨的叙述. 因此，很多定理都舍掉了证明过程，或者即使进行了证明，也没有把重点放在严谨的叙述而仅仅使读者停留于大概了解的程度. 读者在阅读过程中即使没有完全理解，也不必过份在意而可继续向前. 因为对于这些读者来讲，最重要的是学会怎样把定理做为工具使用，而不是对其证明过程的理解.

本书所有的例题都选自统计学、物理学、运筹学等概率论和随机过程的应用领域. 每一章后面都配有一定的习题并附有解答. 至于某些章节中提出的问题，考虑到只要能够理解问题前面的内容即可解决它们，故没有给出解答.

另外，考虑到这套丛书的针对对象，本书还删去了不少重要内容. 比如，在原稿中曾准备过有关布朗运动 (Brown movement) 和概率微分方程等状态连续的随机过程内容，但最后还是删去了. 至于这是否妥当，只好根据选材范围和程度深浅由读者判断. 希望能够得到批评和指正.

本书在这套丛书编委，东京大学甘利俊一先生的建议下执笔.

最初预定由他和我共著这本书，但因故后来由我一人执笔，使得大家未能领略到他的博识，感到非常遗憾．千叶工业大学山本彬也先生曾审阅原稿，并提出了许多宝贵意见，在此，对他表示深深地感谢．同时，对从这套丛书筹备工作开始自始至终给予我鞭策鼓励的讲谈社芳贺骥先生也表示深深地谢意．

伏见正则

1987 年 2 月于东京

第 1 章 概率空间

1.1 概率空间

掷一粒骰子，观察出现的点数. 毋庸置疑，它有 6 种可能结果，即出现的结果必在 1 到 6 之间，可是我们事先却不能确定将会出现哪一个点数. 一般来讲，把能够事先明确一次试验的所有可能结果，但却不能具体确定将会出现哪一种结果的试验称做 **随机试验**, 简称 **试验** (experiment, trial). 在随机试验中，每一个可能出现的结果称做 **样本点** (sample point), 而样本点的全体 Ω 称做 **样本空间** (sample space). 刚才所举的掷骰子的例子中，试验的样本空间 $\Omega = \{1,2,3,4,5,6\}$.

掷骰子，根据不同的目的和不同的情况，有时我们希望它出现偶数点或奇数点，有时我们希望它出现不小于 5 的点数. 这类我们所关心的事件，称做 **随机事件**, 简称 **事件** (event). 从数学的角度，事件也可描述成为 **样本空间的部分集合**. 为了方便，在本书的以后部分，我们将把事件和部分集合当同一名词使用. 刚才例子中的两个随机事件分别相当于部分集合 $A_1 = \{2,4,6\}$, $A_2 = \{1,3,5\}$, $A_3 = \{5,6\}$.

在一次随机试验中，我们观察到了一个样本点 $\omega \in \Omega$. 如果 ω 属于部分集合 $A(\omega \in A)$，那么我们就称发生了事件 A. 比如在掷骰子的例子中，如果出现了 $\omega = 5$, 我们就称发生了事件 A_2 和 A_3.

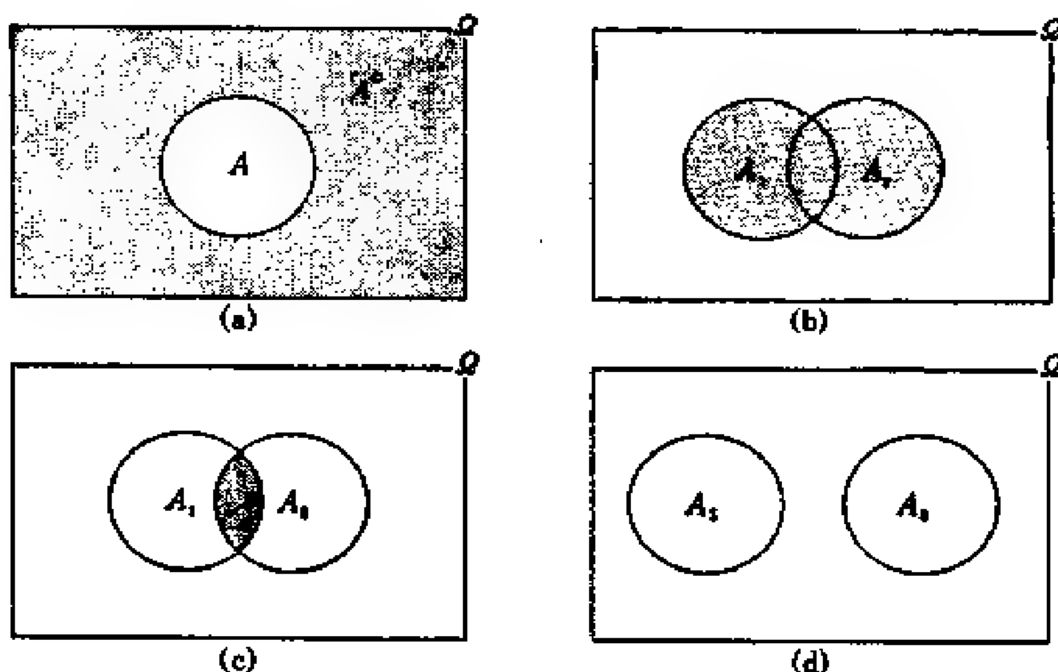


图 1-1 (a) 对立事件 A^c (b) 和事件 $A_1 \cup A_2$
(c) 积事件 $A_1 \cap A_2$ (d) 互不相容事件 A_1 与 A_2

对于任意事件 A 来讲，其**对立事件** (complementary event) A^c 是指当且仅当事件 A 没有发生时发生的事件 (如图 1-1). 按集合论的表示方法，它记做

$$A^c = \{\omega \in \Omega : \omega \notin A\} \tag{1.1}$$

对于任意两个事件 A_1 与 A_2, 其**和** (sum of events) $A_1 \cup A_2$ 是指事件 A_1 与 A_2 中至少有一个发生的事件，记做

$$A_1 \cup A_2 = \{\omega \in \Omega : \omega \in A_1 \quad \text{或} \quad \omega \in A_2\} \tag{1.2}$$

对于任意两个事件 A_1 与 A_2, 其**积** (intersection of events) $A_1 \cap A_2$ 是指事件 A_1 与事件 A_2 同时发生的事件，记做

$$A_1 \cap A_2 = \{\omega \in \Omega : \omega \in A_1 \quad \text{且} \quad \omega \in A_2\} \tag{1.3}$$

特别地，当 $A_1 \cap A_2 = \phi$ 的时候，即集合 A_1 与 A_2 互质的时候，我们称事件 A_1 与事件 A_2 **互不相容** (disjoint events).

上面所讲的和与积的定义还可推广到 3 个及 3 个以上事件的情况. 如果存在 3 个及 3 个以上事件，并且其中任意两个事件互不相容，那么我们称这些事件 **互不相容** (mutually disjoint).

一般来讲，我们所关心的事件将随着目的和场合的不同而不同，就是在同一样本空间 $\mathcal{A}$ 里，各种各样的事件群都可能成为被考察的对象. 在概率论中，针对某一具体目的的事件群 $\mathcal{A}$, 我们要求它满足以下条件：

[B1] Ω 包含在 $\mathcal{A}$ 中.

[B2] 如果事件 A 包含在 $\mathcal{A}$ 中，其对立事件 A^c 也包含在 $\mathcal{A}$ 中.

[B3] 如果事件 $A_1, A_2, \cdots$ 包含在 $\mathcal{A}$ 中，其和 $A_1 \cup A_2 \cup \cdots$ $(= \bigcup_{i=1}^{\infty} A_i)$ 也包含在 $\mathcal{A}$ 中.

同时满足上面 3 个条件的 $\mathcal{A}$ 称做 **波雷尔域** (Borel field). 根据上面的条件，我们还容易推知以下条件.

[B4] 空集合 ϕ 包含在 $\mathcal{A}$ 中.

证明 因为 $\phi = \Omega^c$, 故从 [B1] 及 [B2] 可推出 [B4].

[B5] 如果事件 $A_1, A_2, \cdots$ 包含在 $\mathcal{A}$ 中，其的积 $A_1 \cap A_2 \cap \cdots$ $(= \bigcap_{i=1}^{\infty} A_i)$ 也包含在 $\mathcal{A}$ 中.

证明 由集合论中的 **德 · 摩尔根定律**(De Morgan Law) 法则，

$$(\bigcap_{i=1}^{\infty} A_i)^c = \bigcup_{i=1}^{\infty} A_i^c \tag{1.4}$$

及 [B2], [B3] 可以推出 [B5].

确定好我们所关心的作为事件集合 $\mathcal{A}$ 的波雷尔域以后，再去考察包含在 $\mathcal{A}$ 中的种种事件所发生的 **概率**(有时被称做 **概率测**

度, probability measure). 习惯上，我们将事件 A 发生的概率记为 $P(A)$. 概率 $P(\cdot)$ 具有以下性质：

[P1] 对于任意事件 A, 其概率为介于 0 到 1 之间的实数. 即

$$0 \leq P(A) \leq 1. \tag{1.5}$$

[P2] 样本空间全体 Ω 的概率为 1, 即

$$P(\Omega) = 1 \tag{1.6}$$

[P3] 如果事件 $A_1, A_2, \cdots$ 两两互不相容，则有

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{1.7}$$

性质 [P3] 称为概率的 **完全可加性** (complete additivity).

这一节里，我们首先考察了试验可能发生的所有结果的集合 Ω, 接着导入了作为 Ω 部分集合的集合并满足性质 [B1]-[B3] 的波雷尔域 $\mathcal{A}$, 最后定义了作为对应于 $\mathcal{A}$ 的元素的实函数并满足性质 [P1]-[P3] 的概率 $P(\cdot)$. 在概率论中，把 Ω, $\mathcal{A}$ 及 P 三者的结合体 $(\Omega, \mathcal{A}, \mathcal{P})$ 称做 **概率空间** (probability space).

1.2 概率的基本公式

概率问题，很少能够 (或者不可能) 做到事先举出波雷尔域中所有部分集合 (事件) 的概率. 一般来讲，我们通过基本事件的概率算出其它事件的概率. 这时，需要用到各种各样从概率基本性质 [P1]-[P3] 推出的公式. 下面列举一些这类公式.

[P4] $P(\phi) = 0$ (1.8)

[P5] 如果 $A_1, A_2, \cdots, A_n$ 两两互不相容，则有

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \tag{1.9}$$

[P6] 对于任意 A,

$$P(A^c) = 1 - P(A) \tag{1.10}$$

证明 令 $A_1 = \Omega$, $A_2 = A_3 = \cdots = \phi$, 再利用 [P3] 和 [P2] 便可推出 [P4]. 又令事件 $A_1, A_2, \cdots A_n$ 两两互不相容, $A_{n+1} = A_{n+2} = \cdots = \phi$, 利用 [P3] 和 [P4] 便可推出 [P5]. 最后令 $n = 2, A_1 = A, A_2 = A^c$, 然后利用 [P2] 及 [P5] 便可推出 [P6].

[P7] $P(\bigcup_i A_i) = 1 - P(\bigcap_i A_i^c)$. (1.11)

证明 令 $A = \bigcup_i A_i$ (有限或者无限个事件之和), 然后利用 [P6] 及德·摩尔根定律 (1.4) 便可推出 [P7].

[P7] 表示 $A_1, A_2, \cdots$ 中至少有一个事件发生的概率等于 1 减去没有任何一个事件发生的概率 (如图 1-2).

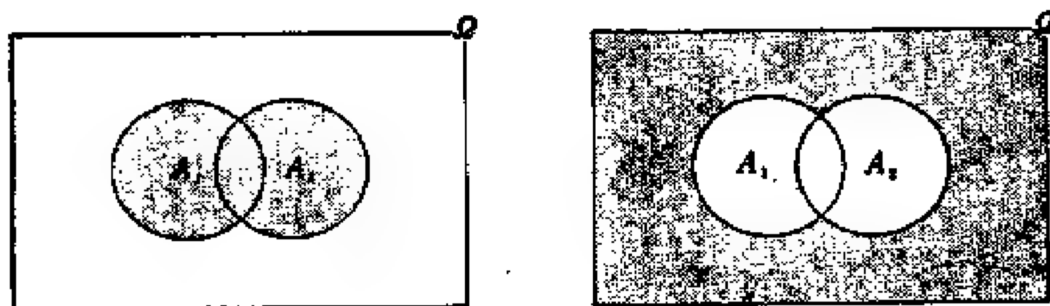


图 1-2 和事件 $A_1 \cup A_2$ 与其对立事件 $(A_1 \bigcup A_2)^c = A_1^c \cap A_2^c$

计算复杂事件的概率问题时，通常把该事件分解成便于计算概率的简单事件进行考虑. 这种思维方法基于下面的性质 [P8].

[P8] 对于两两互不相容的事件 $B_1, B_2, \cdots$, 如果有 $\cup_i B_i = \Omega$.

则对于任意事件 A,

$$P(A)=\sum_i P(A\cap B_i) \tag{1.12}$$

证明 根据集合论中的分配定律 (如图 1-3),

$$A=A\cap\Omega=A\cap(\bigcup_i B_i)=\bigcup_i(A\cap B_i)$$

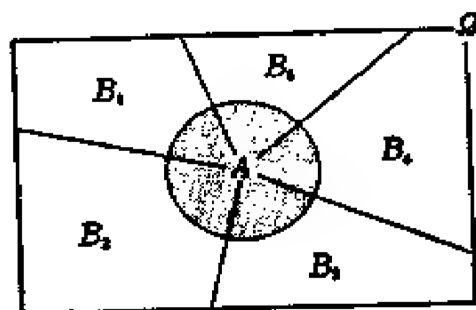


图 1-3 $A=\bigcup_{i=1}^5(A\cap B_i)$

因为序列 $B_i(i=1,2,\cdots)$ 互不相容, 故序列 $A\cap B_i(i=1,2,\cdots)$ 也互不相容. 故由 [P5] 得到式 (1.12).

如果 A_1,A_2 互不相容, 则由 [P5]

$$P(A_1\cup A_2)=P(A_1)+P(A_2).$$

[P9] 如果 A_1,A_2 不一定互不相容, 则

$$P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1\cap A_2) \tag{1.13}$$

概率论中, 我们称它为 **加法公式** (addition law). 由图 1-4 可知, 如果仅仅通过将 $P(A_1)$ 与 $P(A_2)$ 相加来求 $P(A_1\cup A_2)$, 就会多

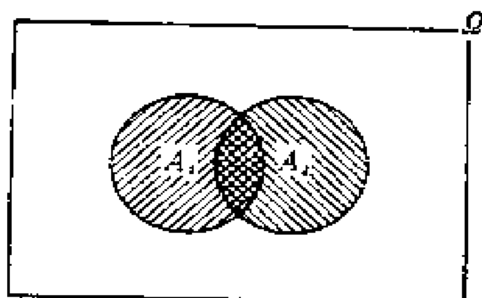


图 1-4 加法公式说明图 (1)

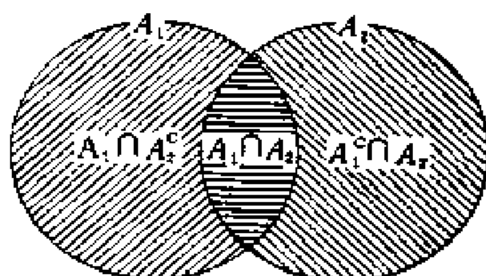


图 1-5 加法公式说明图 (2)

出对应于 $A_1 \cap A_2$ 的那部分概率，所以我们又通过减去 $P(A_1 \cap A_2)$ 来加以调整.

证明 形式上可以证明如下. 由图 1-5,

$$A_1 \cup A_2 = (A_1 \cap A_2^c) \cup (A_1^c \cap A_2) \cup (A_1 \cap A_2)$$

因为右边 3 个括号内的集合互不相容，根据 [P5]

$$P(A_1 \cup A_2) = P(A_1 \cap A_2^c) + P(A_1^c \cap A_2) + P(A_1 \cap A_2) \tag{1.14}$$

又由 [P8],

$$P(A_1 \cap A_2^c) = P(A_1) - P(A_1 \cap A_2)$$
$$P(A_1^c \cap A_2) = P(A_2) - P(A_1 \cap A_2)$$

最后将上式代入 (1.14), 便可推出 [P9].

加法公式还可推广到如下 3 个事件的情况.

[P10] $P(A_1 \cup A_2 \cup A_3) =$

$$\begin{aligned}&P(A_1) + P(A_2) + P(A_3)\\&- \{P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_3 \cap A_1)\}\\&+ P(A_1 \cap A_2 \cap A_3)\end{aligned} \tag{1.15}$$

仿 [P9], 我们通过图 1-6 来理解这个公式. A_1 为右上阴影, A_2 为右下阴影, A_3 为水平阴影, 求 $A_1 \cup A_2 \cup A_3$ 的概率, 如果仅将 $P(A_1), P(A_2), P(A_3)$ 相加, 那么二重及三重阴影部分将分别被加上二次和三次. 这样, 首先为了修正被加上的二重阴影部分而从上式减去 $P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_3 \cap A_1)$. 这时, 又多减去了三倍的三重阴影部分的概率, 即等于根本没加上这一部分. 于是, 最后加上这一部分概率 $P(A_1 \cap A_2 \cap A_3)$ 即可了.

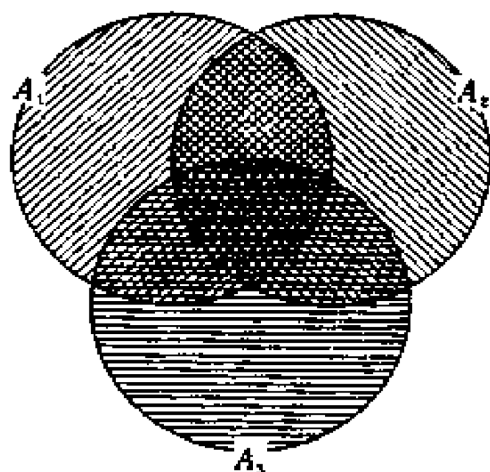


图 1-6　关于三个事件的加法公式说明图

一般来讲, 对于四个及四个以上的事件, 同样具有如下的加法公式.

[P11] 如果 $S_m = \sum_{i_1 < i_2 < \cdots i_m} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m})$, 则

$$P(\bigcup_{i=1}^{n} A_i) = S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n-1} S_n \tag{1.16}$$

这里, S_m 右边的 Σ 算子把从 n 个整数 $1, 2, \cdots, n$ 中取出 m 个相

异整数的所有组合进行相加.[1]

例 1-1　一个晚会，共有 n 位参加者. 每人来时带了一件礼物，主办者把这些礼物收集在一起，并在晚会结束参加者回去的时候把它们随机地发给每个人. 问：至少有一个人带回自己带来礼物的概率有多大？ [这个问题，被称为 **门摩尔问题**(Monmort problem)，或称为 **偶然的一致性问题**(coincidences problem).]

首先，把从 1 到 n 的番号标给每个参加者，并把番号为 i 的参加者带回自己带来礼物的事件记为 A_i. 那么 A_i 发生的概率，相当于把从 1 到 n 的数字分别标在 n 张牌上，随便洗好牌后从左到右一张一张发牌时第 i 次出现数字 i 的概率. 因为，随便洗好牌并把这 n 张牌排成一列时的数排列方法共有 $n!$ 种，且其中每一种可能出现的概率均为 $1/n!$. 又因为 A_i 是第 i 次出现数字 i 的事件，所以用不着考虑剩下的 n-1 张牌将以什么样的顺序出现，即

$$P(A_i) = (n-1)!/n!.$$

同理，

$$P(A_{i_1} \cap A_{i_2}) = (n-2)!/n!$$
$$P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = (n-3)!/n!$$
$$\cdots\cdots\cdots$$
$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = (n-m)!/n!$$

---

[1] 这种组合，共有

$$\binom{n}{m} = n(n-1)\cdots(n-m+1)/m! = \frac{n!}{(n-m)!m!}$$

$\binom{n}{m}$ 称为二项系数，也可记为 C_n^m.

因而

$$S_m = \binom{n}{m}\frac{(n-m)!}{n!} = \frac{1}{m!}$$

$$P_1 = P(\bigcup_{i=1}^{n} A_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}$$

下面给出了 $n = 3, 4, 5, 6, 7$ 时 P_1 的值.

n =	3	4	5	6	7
P1=	.66667	.62500	.63333	.63194	.63214

由于 e^x 在 $x = 0$ 附近的泰勒级数 (Taylor series)(即麦克劳林级数， Maclaurin series) 为

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

可以想象，当 n 不是很小的时候， P_1 与 $1 - e^{-1}$ 将会非常接近. 事实上，当 $n = 7$ 的时候， $1 - e^{-1} = 0.63212\cdots$ 已经非常逼近 P_1. P_1 与 n 基本上无关这一事实有很深的现实意义.

1.3 概率的连续性

有时候，我们会考虑由无限多个事件组成的序列的概率极限问题. 这时，下面的性质非常重要.

[P12] 对于无限序列 $A_1, A_2, A_3, \cdots$, 假定存在着单调包含关系 $A_1 \subset A_2 \subset A_3 \cdots$. 如果令 $A = \bigcup_{i=1}^{\infty} A_i$, 则

$$P(A) = \lim_{i\to\infty} P(A_i) \tag{1.17}$$

[P13] 对于无限序列 $A_1, A_2, A_3 \cdots$, 假定存在着单调包含关系 $A_1 \supset A_2 \supset A_3 \supset \cdots$, 如果令 $A = \bigcap_{i=1}^{\infty} A_i$, 则有

$$P(A) = \lim_{i\to\infty} P(A_i) \tag{1.18}$$

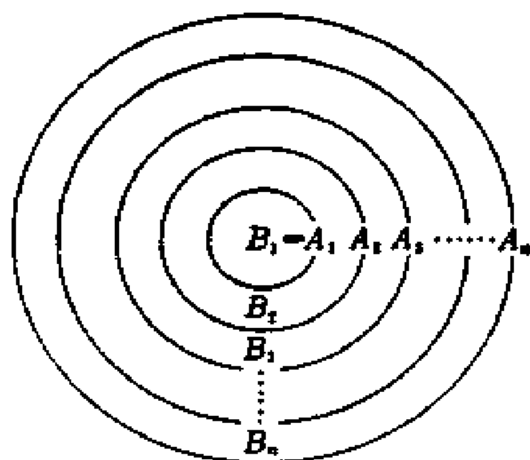


图 1-7　概率的连续性 [P12] 示意图

[P12] 和 [P13] 在某种意义上表明了概率的连续性. 即如果单调递增或单调递减事件序列 $A_1, A_2, \cdots$ 的极限为 A 的时候, $P(A_1), P(A_2), \cdots$ 的极限存在, 且它等于 P(A).

证明　证明 [P12]. 令 $B_1 = A_1, B_2 = A_2 \cap A_1^c, B_3 = A_3 \cap A_2^c, \cdots$ (如图 1-7), 因为 $B_1, B_2, B_3, \cdots$ 互不相容, $\bigcup_{i=1}^{n} B_i = A_n$, $\bigcup_{i=1}^{\infty} B_i = A$. 再由 [P5], $P(A_n) = \sum_{i=1}^{n} P(B_i)$. 当 $n \to \infty$ 的时候, $\lim_{n\to\infty} P(A_n) = \sum_{i=1}^{\infty} P(B_i)$. 最后根据 [P3] (完全可加性), 有

$$\sum_{i=1}^{\infty} P(B_i) = P(\bigcup_{i=1}^{\infty} B_i) = P(A)$$

问题　试证 [P13]. [提示: $A_1^c, A_2^c, \cdots$ 为单调递增事件的序列, 并利用 [P12] 及 [P13].]

1.4 条件概率

例题 1-1 在一次抽奖会上，抽选箱里装有 3 个红球和 7 个白球. 摇一次抽选箱，就会掉出来 1 个球. 如果这个球是红的就中奖，白的就没有中奖. 现在假定掉出来的球不再放回抽选箱中.[1] 这时最初摇奖的人中奖的概率为 3/10. 问：

(1) 第一次摇奖的人中奖的情况下，第二次摇奖的人中奖的概率是多少？

(2) 第一次摇奖的人没有中奖的情况下，第二次摇奖的人中奖的概率是多少？

(1) 第一次摇奖的人中奖以后，抽选箱中还剩下 2 个红球和 7 个白球. 因而，接着出现红球的概率为 $\frac{2}{2+7}=\frac{2}{9}(<\frac{3}{10})$

(2) 第一次摇奖的人没有中奖以后，抽选箱中还剩下 3 个红球和 6 个白球. 因而，接着出现红球的概率为 $\frac{3}{3+6}=\frac{3}{9}(>\frac{3}{10})$

这个例子说明，最初的试验结果一般会影响第二次试验结果的概率. 如果假定最初试验结果发生了事件 B, 那么在第二次试验中事件 A 发生的概率就可记为 $P(A|B)$, 并称它为在事件 B 发生的条件下事件 A 发生的 **条件概率** (conditional probability).

定义

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \tag{1.19}$$

这个式子只限于 $P(B)>0$ 时的情况. 当 $P(B)=0$ 的时候,

[1] 这种抽选方法，叫做 **无放回抽样** (sampling without replacement). 与此相反，如果掉出来的球重新放回到抽选箱中去，这种抽选方法，叫做 **有放回抽样** (sampling with replacement).

我们不定义条件概率.

例题 1-2 以日本具有两个孩子的家庭为对象，分别考察家庭中孩子们的性别. 如果将每个家庭的第一个孩子和第二个孩子区别开来，那么共有四种组合 Ω ={男男，男女，女男，女女}. 现在随机地观察一个家庭，那么它是 Ω 中任意一个情况的概率为 1/4. 现在我们已经知道这个家庭中有一个男孩 (称做事件 B), 问这个家庭中剩下的一个孩子也是男孩 (称做事件 A) 的概率.

因为 B ={男男，男女，女男}, $A\cap B$ ={男男}, 所以

$$P(A|B) = P(A\cap B)/P(B) = \frac{1}{4}/\frac{3}{4} = \frac{1}{3}$$

下面列举条件概率的有关性质和公式.

[C1] 固定 B, 然后将 $P(A|B) = P(A)$ 看做 A 的函数，那么 $P(A)$ 就是对应于样本空间 Ω 的波雷尔域 $\mathcal{A}$ 的概率. 这里，第 3 节中讲过的公式均可适用.

[C2] 式 (1.19) 一般通过 $P(A\cap B)$ 计算 $P(A|B)$, 但根据具体情况，有时还可在事先已知 $P(A|B)$ 的情况下，用它计算 $P(A\cap B)$. 这时，式 (1.19) 改写成为

$$P(A\cap B) = P(B)P(A|B) \tag{1.20}$$

并把它称做 **乘法公式** (multiplication law).

[C3] 如果事件 $B_1, B_2, \cdots$ 两两互不相容，且满足 $\bigcap_i B_i = \Omega$, 则对于任意 A, 都有

$$P(A) = \sum_i P(B_i)P(A|B_i) \tag{1.21}$$

这就是 **全概率公式** (total probability law). 特别地，当 $P(A) > 0$

的时候，

$$P(B_j|A)=\frac{P(B_j)P(A|B_j)}{\sum\limits_i P(B_i)P(A|B_i)} \tag{1.22}$$

这称为贝叶斯 (Bayes formula) 公式.

证明 由概率的性质，从 [P8] 及 [C2] 推出式 (1.21). 另外，由于

$$P(B_j|A)=\frac{P(B_j\cap A)}{P(A)}=\frac{P(B_j)P(A|B_j)}{P(A)}$$

如果将式 (1.21) 代入到分母中去，则又可推出式 (1.22).

例 1-2 下面考虑在例题 1-1 中第一次摇奖的人与第二次摇奖的人哪一种情况更有利. 第一次摇奖的人中奖的概率为 3/10, 没有中奖的概率为 7/10. 第二次摇奖的人中奖的概率，由例题 1-1 的结果及全概率公式，可知为

$$\frac{3}{10}\times\frac{2}{9}+\frac{7}{10}\times\frac{3}{9}=\frac{3}{10}$$

显然，中奖的概率与摇奖的先后顺序没关系. 这在一般情况下也会成立. 比如在选举的时候，为了选出候选人的竞选顺序而事先进行抽签仪式，而抽签仪式中的抽签顺序，人们又往往通过抽签方式来决定. 这种做法，从概率论的角度上讲，是没有任何意义的.

例 1-3 某工厂为了制造某种设备而从三个公司 B_1,B_2,B_3 购买同一种零部件，购买的数量之比分别为 50%, 30% 和 20%. 现在假定安装到一台设备中去的这种零部件由各公司制造的可能性也分别为 50%,30% 和 20%, 另外又假定从各公司购买的零部件在使用一年以后发生故障的概率分别为 0.015, 0.025 和 0.035. 如果在使用一年以后有这么一种零部件发生了故障 (此事件称做 A), 问这零部件由 B_1,B_2,B_3 制造的概率为多大？

因为 $P(B_1)=0.50$, $P(B_2)=0.30$, $P(B_3)=0.20$, $P(A|B_1)=0.015$, $P(A|B_2)=0.025$, $P(A|B_3)=0.035$,

$$\sum_{i=1}^{3} P(B_i)P(A|B_i) = 0.0075+0.0075+0.0070 = 0.022$$

再根据贝叶斯公式,

$$P(B_1|A) = P(B_2|A) = \frac{0.0075}{0.022} \approx 0.34$$
$$P(B_3|A) = \frac{0.0070}{0.022} \approx 0.32$$

1.5 独立性

从第 4 节所讲的内容，我们知道一般来讲，概率 $P(A)$ 并不等于条件概率 $P(A|B)$. 也就是说，事件 B 的发生将会影响事件 A 的发生. 但是根据具体情况的不同，有时也会有

$$P(A|B) = P(A) \tag{1.23}$$

即事件 A 的发生难易程度根本不受事件 B 的影响. 这时，我们称事件 A 与事件 B **相互独立** (mutually independent). 根据条件概率的定义式 (1.19), 式 (1.23) 在 $P(B)>0$ 的情况下

$$P(A\cap B) = P(A)P(B) \tag{1.24}$$

因为式 (1.24) 在 $P(B)=0$ 的时候也可适用，另外此式中 A 与 B 具有对称的形式，故从独立性的角度上讲，式 (1.24) 具有更大的意义.

定义 所谓两个事件 A, B 相互独立，是指满足公式 (1.24).

将式 (1.24) 与式 (1.20)(乘法公式) 进行比较，就会注意到它是

A 和 B 相互独立时的乘法公式. 对于事件 A, B, C 来讲, 如果满足等式

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(B \cap C) &= P(B)P(C) \\ P(C \cap A) &= P(C)P(A) \end{aligned} \tag{1.25}$$

则根据定义可知, 事件 A 与 B, 事件 B 与 C 及事件 C 与 A 相互独立. 但是, 这时等式

$$P(A \cap B \cap C) = P(A)P(B)P(C) \tag{1.26}$$

却不一定成立. 比如, 我们不能因此而简单地认为 $A \cap B$ 与 C 相互独立. 下面举一个简单的例子说明这个问题.

令 $\Omega = \{1, 2, 3, 4\}$, 各样本点都被赋以 1/4 的概率. 如果 $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$, 则 $P(A) = P(B) = P(C) = 1/2$. 又由于 $A \cap B = B \cap C = C \cap A = \{1\}$, 式 (1.25) 成立. 但是, $P(A \cap B \cap C) = P(\{1\}) = 1/4 \neq P(A)P(B)P(C)$. 事实上 $P(C|A \cap B) = 1$, 即如果 $A \cap B$ 发生, 则事件 C 也一定发生. 可见, 事件 $A \cap B$ 与 C 并非互为独立.

我们通常所指的"独立"概念是指 A, B, C 相互独立, 则它们的任意组合也相互独立, 即如果式 (1.25) 与式 (1.26) 同时成立, 可将事件 A, B, C 定义为相互独立.

更一般地, n 个事件的独立性也可如下定义.

定义 所谓 n 个事件 $A_1, A_2, \cdots, A_n$ 相互独立, 是指对于其中任意事件 $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$, 总有等式

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}) \tag{1.27}$$

成立.

问题 如果事件 A, B, C 互为独立，试证明 $A \cup B$ 与 C 也互为独立.

问题 如果事件 $A_1, A_2, \cdots, A_n$ 相互独立，试证将它们分成 m 组 $(m < n)$, 并对各组事件施以 $\cup$ (和事件), $\cap$(积事件), ^{c} (对立事件) 演算以后得到的事件序列 $B_1, B_2, \cdots, B_m$ 也相互独立.

从**事件的独立性** (independence of events), 我们可以定义试验的独立性，即如果两次或两次以上试验产生的任意事件相互独立，则称这些试验也相互独立.

习 题 1

1. 事先准备好 1 角，5 分和 1 分硬币各一枚，然后往上抛这些硬币，并纪录这些硬币是国徽这一面朝上 (T), 还是数字这一面朝上 (H). 比如 (H, T, H) 是指 1 角硬币出现了数字这一面朝上，5 分硬币出现了国徽这一面朝上，1 分硬币出现了数字这一面朝上.

(1) 试问这个试验的样本空间.

(2) 将 1 角，5 分和 1 分硬币出现数字这一面朝上的事件分别记为 A, B, C. 试列举出以下事件包含的样本点. A, $B \cup C$, $B \cap C$, $A^c \cap B$, $(A \cup B^c) \cap C$

(3) 假定对于每个硬币来讲，出现数字这一面朝上的概率均为 1/2. 试问 (2) 中各事件发生的概率.

2. 设 A, B, C 为三个事件，试用 A, B, C 的运算关系表示下列各事件：

(1) A, B, C 中至少有一个发生.

(2) A, B, C 中哪一个都不发生.

(3) A, B, C 中正好有两个发生.

(4) A, B, C 中至少有两个发生.

(5) A, B 中只有一个发生, C 不发生.

3. 设一台计算机在某一星期内发生故障的概率为 p. 试求某一星期内同时使用 n 台计算机的时候, 至少有一台不会发生故障的概率 p. 又问当 $p = 0.01$ 的时候, 为了使 $P >= 0.9999$, n 至少应为多少?

4. 有一位跳高选手, 每跳 10 次平均有 3 次可以越过 1.5m 以上. 问这位选手在三次以内能够越过这个高度的概率为多少.

5. 试证明三个事件的和的公式 (1.15).

6. ABO 式血型, 可由染色体中一对遗传因子的组合而定, 其中 A,B 为显性, O 为隐性. 如表 1-1 所示, 遗传因子的组合 (遗传因子型) 共有六种, 血型 (表现型) 共有 A,B,AB,O 四种. 表中的百分比是以日本人为对象的各种频率 (选自《平凡社世界大百科事典》). 孩子可从父母那里分别得到一个遗传因子, 且孩子从父亲或母亲的一对遗传因子中得到某一个遗传因子的概率均为 1/2.

表 1-1　ABO 血型的构成与频率 (日本人)

表现型	频率 (%)	遗传因子型	频率 (%)
A 型	39	AA	8
		AO	31
B 型	22	BB	3
		BO	19
AB 型	10	AB	10
O 型	29	OO	29

(1) 试证对于父母血型的任意组合, 生出来的孩子表现出的血型频率 (百分比) 与表 1-2 一致.

表 1-2　ABO 血型的遗传（日本人）

双亲血型的组合	各种子女血型出现的频率（%）			
	O	A	B	AB
O×O	100	0	0	0
O×A	40	60	0	0
O×B	43	0	57	0
O×AB	0	50	50	0
A×A	16	84	0	0
A×B	17	26	23	34
A×AB	0	50	20	30
B×B	19	0	81	0
B×AB	0	22	50	28
AB×AB	0	25	25	50

(2) 如果父母二人的血型一个是 O 型，一个是 A 型，并假定最初生的孩子的血型为 A 型，试问父母二人中血型为 A 型那一位的遗传因子型为 AO 的概率为多少.

(3) 如果父母二人的血型一个是 A 型，一个是 B 型，并假定最初生的孩子的血型为 AB 型，试问父母二人中血型为 A 型那一位的遗传因子型为 AO 的概率为多少. 又问血型为 B 型那一位的遗传因子型为 BB 的概率为多少.

第 2 章 随机变量

2.1 随机变量及其分布函数

例 2-1 某一商店在年末大甩卖中实行了全奖抽选. 摇奖时从摇箱出来的球的颜色分别为红、蓝、黄、绿、白中的一种，对应的中奖额分别为 10000 日元，5000 日元，1000 日元，500 日元和 100 日元. 假定摇箱里装有很多球，且各种颜色球的百分比分别为 0.1%, 0.5%, 2%, 10% 和 87.4%. 这里，我们关心的是中奖额，现把它们记做 X. 由于样本空间 $\Omega = \{$ 红，蓝，黄，绿，白 $\}$, 且 X 可由 Ω 中每个样本点 ω 决定，因而 X 也可看做是 ω 的函数. 对于每个 $\omega(\in \Omega)$, X 的取值 $X(\omega)$ 及其概率可写成表 2-1 的形式 ($X(\omega)$ 的单位为日元).

表 2-1 随机变量例

ω	红	蓝	黄	绿	白
$X(\omega)$	10000	5000	1000	500	100
概率	0.001	0.005	0.02	0.1	0.874

一般地，对于任意概率空间 $(\Omega, \mathcal{A}, P)$, 同样可以考虑这么一个函数 X, 即对应于每个 $\omega(\in \Omega)$, 都有一个与其对应的实数值 $X(\omega)$. 值得注意的是，由于我们还要关心 X 取每个实数值的概率，所以只有那些通过 P 能够算出其概率的 X 才是我们的研究对象. 具备这个条件的 X 被称做 **随机变量** (random variable). 事实上，如

果注意到 P 是对应于包含在 $\mathcal{A}$ 中的集合定义的，那么我们自然而然就会得到如下定义.

定义 定义在概率空间 $(\Omega, \mathcal{A}, P)$ 上的实函数 $X(\omega)(\omega \in \Omega)$, 如果对于所有的实数值 x 都有

$$\{\omega : X(\omega) \le x)\} \in \mathcal{A} \tag{2.1}$$

则称 $X(\omega)$ 为随机变量.

对于所有 x, 如果已知式 (2.1) 中集合 $\{\omega : X(\omega) \le x\}$ 对应的概率，那么与随机变量 $X(\omega)$ 有关的各个事件的概率能够简单地通过计算求出而不用深入到概率空间 $(\Omega, \mathcal{A}, P)$ 中去. 一般地，我们可以从

$$F_X(x) = P(\{\omega : X(x) \le x\}) \qquad (-\infty < x < \infty) \tag{2.2}$$

出发去讨论 $X(\omega)$. 这个函数，被称做**随机变量 X 的分布函数**. 在不致引起误解的情况下，我们可将 X 省去而把它写成 $F(x)$ 的形式. 此外，我们还可将 $P(\{\omega : X(\omega) \le x\})$ 简写成 $P(X \le x)$.

分布函数 $F(x)$ 具有如下性质.

(1) 单调不减性：如果 $a < b$, 则 $F(a) \le F(b)$. (2.3)

(2) 右连续： $\lim\limits_{x \to a+0} F(x) = F(a)$. (2.4)

(3) $\lim\limits_{x \to -\infty} F(x) = 0$. (2.5)

(4) $\lim\limits_{x \to \infty} F(x) = 1$. (2.6)

证明 (1) 令 $A = \{\omega : X(\omega) \le a\}$, $B = \{\omega : X(\omega) \le b\}$. 则

$$\begin{aligned} F(b) - F(a) &= P(B) - P(A) \\ &= P(B \cap A^c) \\ &= P(\{\omega : a < X(\omega) \le b\}) \ge 0 \end{aligned}$$

(2) 在 a 的右侧，任取一个单调数列 $x_1 > x_2 > \cdots$, 那么

$$\begin{aligned}\lim_{n\to\infty} F(x_n) &= \lim_{n\to\infty} P(\{\omega : X(\omega) \le x_n\}) \\ &= P(\lim_{n\to\infty}\{\omega : X(\omega) \le x_n\}) \\ &= P(\{\omega : X(\omega) \le a\}) \\ &= F(a)\end{aligned}$$

上式的第二个等号可由概率的连续性 [P13] 推知.

(3) 在 (2) 中取一个满足 $\lim_{n\to\infty} x_n = \infty$ 的数列即可.

(4) 取一个单调数列 $x_1 < x_2 < \cdots$, $\lim_{n\to\infty} x_n = \infty$, 然后仿 (2) 证明 (这时要用到概率的连续性 [P12]).

分布函数，并非一定要连续 (即它虽为右连续，但却不一定是左连续), 读者很容易从下节所述各分布函数的例子中看出这一点.

2.2 离散型分布和连续型分布

2.2.1 离散型分布

如果随机变量 X 的所有可能取值是有限或无限可列的 (在有限区间内则是有限的), 则 X 呈现出 **离散型概率分布**(或称 **离散分布**) (discrete distribution). 假定 X 取 $V = \{v_0, v_1, v_2, \cdots\}$ 中任意一个值 v_k, 则其概率可由 X 的分布函数 $F(x)$ 算出，即可选定一个充分小的 $\varepsilon > 0$, 使得在区间 $[v_k - \varepsilon, v_k]$ 中仅包含 v_k. 这时

$$f(v_k) \equiv P(X = v_k) = F(v_k) - F(v_k - \varepsilon) \tag{2.7}$$

我们称 $f(v)$ 为这个分布的 **概率分布** (probability function). 这里要注意，分布函数是对所有实数值定义的，而概率分布则是对属于 V 的实数值定义的 (或可认为在其它实数值处 $f = 0$).

式 (2.7) 是从分布函数求出概率分布的公式. 相反地, 我们也可从概率分布求出分布函数. 即

$$F(x) = \sum_{v_k \leq x} f(v_k) \tag{2.8}$$

右边的和对应于所有满足 $v_k \leq x$ 的 $f(v_k)$. 可见, 离散分布既可由分布函数也可由概率分布给出.

分布函数满足第 1 节叙述过的性质 (1)-(4), 概率分布满足以下两个性质.

(i) 对于所有 k, $f(v_k) \geq 0$. (2.9)

(ii) $\sum_k f(v_k) = 1$. (2.10)

在应用中, 离散分布的取值 $v_0, v_1, v_2, \cdots$ 常常等间隔地排列在数轴上. 这种分布, 称做 **格状分布** (lattice distribution). 一般地, 我们会经常用到取值为非负整数的格状分布. 下面举几个例子.

(1) **0 - 1 分布** (two point distribution) $\mathrm{B}(1;p)$ $(0 < p < 1)$

随机变量只取 0 和 1 两个值, 且 p 为 0 到 1 之间的任意整数, 即

$$\begin{aligned} &V = \{0, 1\} \\ &f(0) = 1 - p, f(1) = p \end{aligned} \tag{2.11}$$

赌博中的输与赢, 抽签的中与不中, 设备质量的好与坏, 舆论调查的赞成与反对等都可看做 0 - 1 分布. 服从 0 - 1 分布的试验称做 **伯努利试验** (Bernoulli trial).

(2) **二项分布** (binomial distribution) $\mathrm{B}(n;p)$ $(0 < p < 1)$

$$\begin{aligned} &V = \{0, 1, 2, \cdots, n\} \\ &f(k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \end{aligned} \tag{2.12}$$

抽出 n 张奖券，每张都服从 0-1 分布，则有 k 张中奖的概率为 $f(k)$ (设中奖率为 p, 即平均 $1/p$ 张奖券中有一张奖券中奖), 即当进行 n 次伯努利试验的时候，我们所关心的结果 (赌赢，中奖，质量好，赞成) 出现 k 次的概率是 $f(k)$.

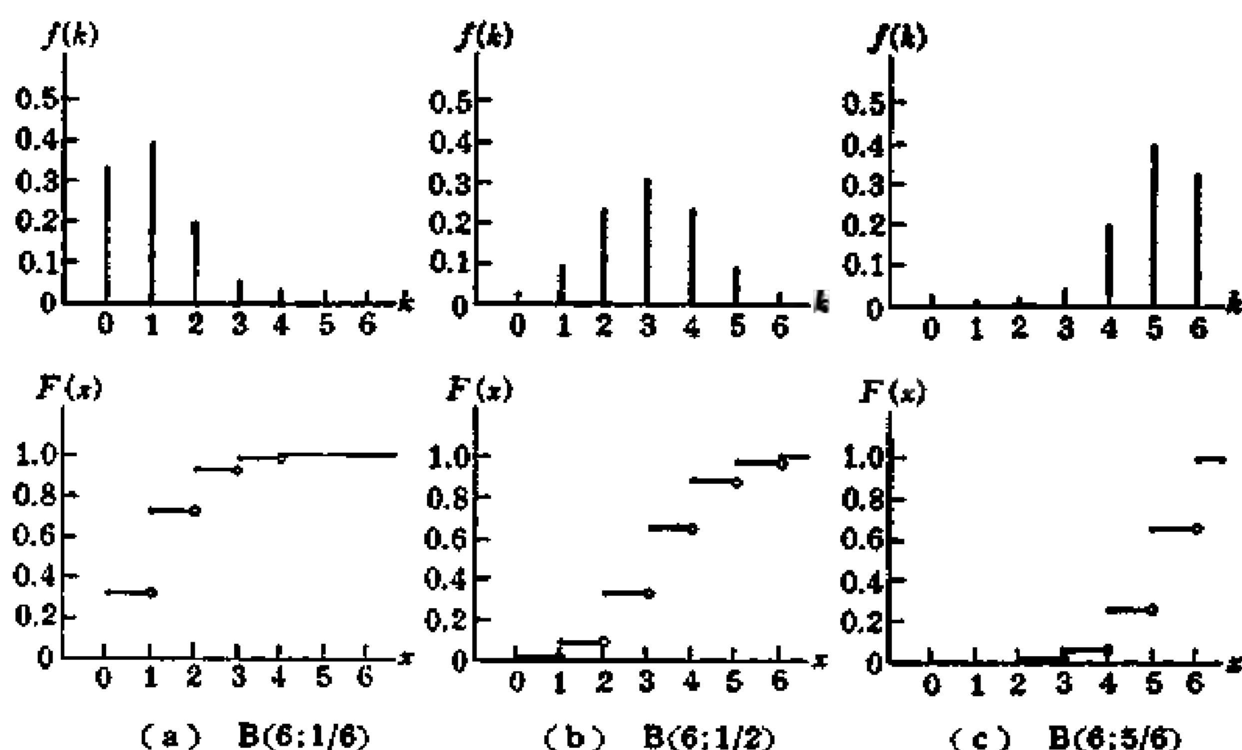


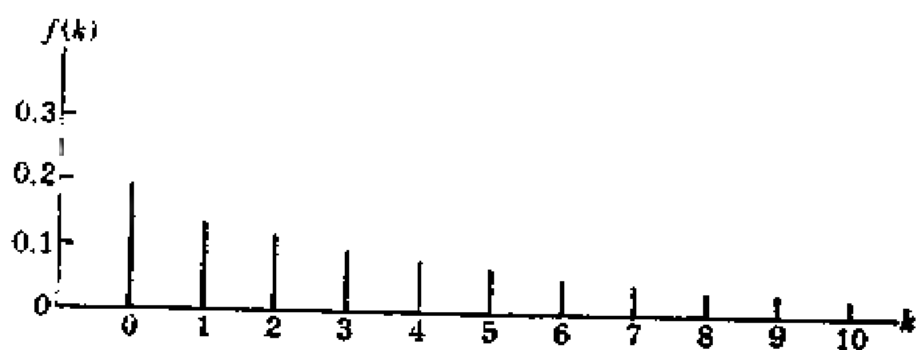
图 2-1　二项分布的概率分布和分布函数.

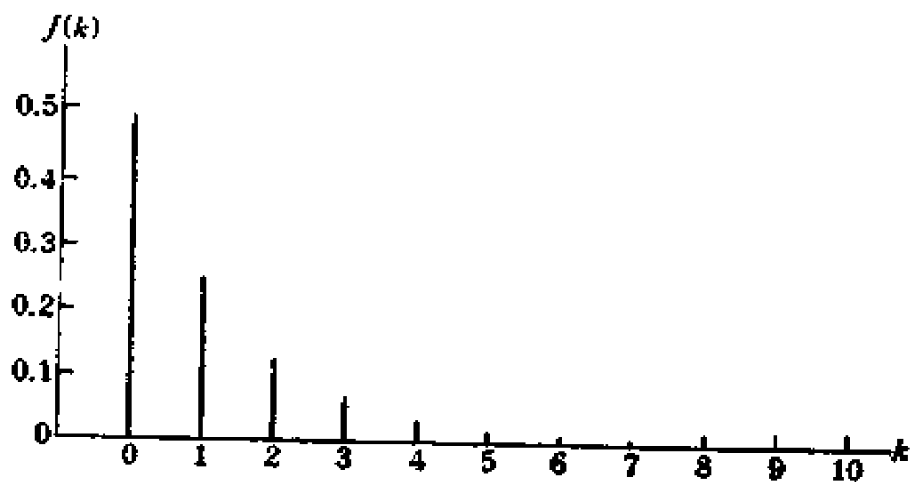
上图：概率分布. 下图：分布函数

(3) **几何分布** (geometric distribution) $\mathrm{Ge}(p)$ $(0 < p < 1)$

$$\begin{aligned} &V = \{0, 1, 2, \cdots\} \\ &f(k) = (1-p)^k p \end{aligned} \tag{2.13}$$

在 (1), (2) 的奖券例子中，一张一张地抽奖，直到首次出现中奖奖券为止. 这时，没有中奖的奖券个数是 n 的概率为 $f(k)$.





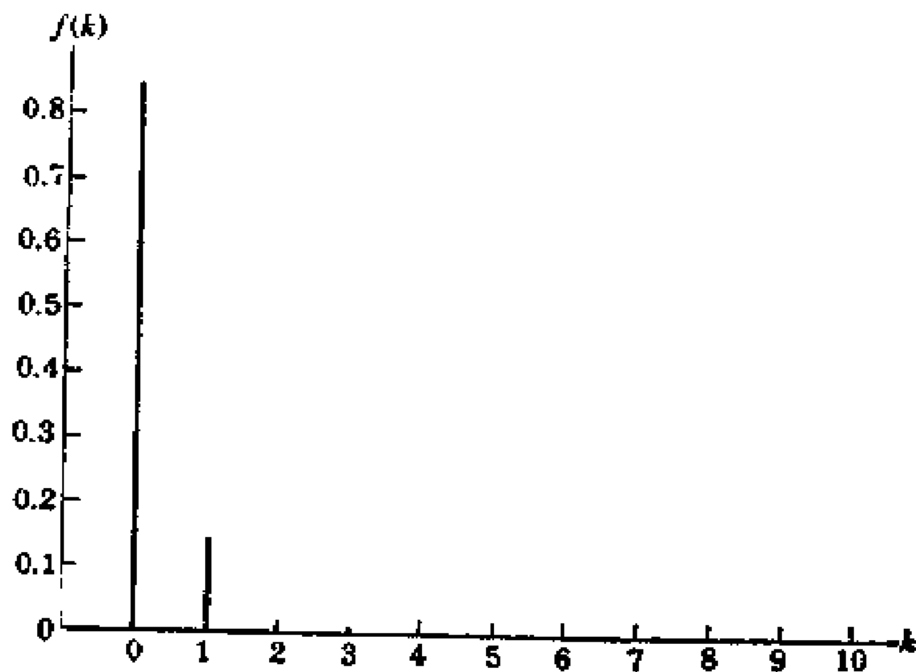


图 2-2 几何分布的概率分布

(4) 泊松分布 (Poisson distribution) Po(λ)

$$
\begin{aligned}
&V = \{0, 1, 2, \cdots\} \\
&f(k) = e^{-\lambda}\frac{\lambda^k}{k!}
\end{aligned}
\tag{2.14}
$$

人们从日常生活中发现，偶然的现象发生的次数近似地服从泊松分布. 例如一定时间内电话交换台收到的电话呼唤次数，一定时间内发生裂变的放射性物质的原子个数，某区域在一个月内发生的交通事故次数等等. 关于它，将在第 5 章做详细介绍.

二项分布与泊松分布的关系

当二项分布 B(n;p) 的 n 很大，p 很小的时候，概率分布 $f(k)$ 非常地接近泊松分布的概率分布. 下面我们来说明这一点. 将 $p = \lambda/n$ 代入二项分布的 $f(k)$，则

$$
\begin{aligned}
f(k) &= \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k} \\
&= \frac{\lambda^k}{k!}\{(1-\frac{1}{n})(1-\frac{2}{n})\cdots \\
&\qquad (1-\frac{k-1}{n})(1-\frac{\lambda}{n})^{-k}\}(1-\frac{\lambda}{n})^n
\end{aligned}
\tag{2.15}
$$

这里，在保持 λ 和 k 不变的情况下令 $n \to \infty$, 则 { } 内各项趋近于 1, 最后一项趋近于 $e^{-\lambda}$, 因而

$$
\lim_{n\to\infty} f(k) = e^{-\lambda}\frac{\lambda^k}{k!} \tag{2.16}
$$

特别地，当 n 很大的时候，有近似公式

$$
\begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!} \quad (\lambda = np) \tag{2.17}
$$

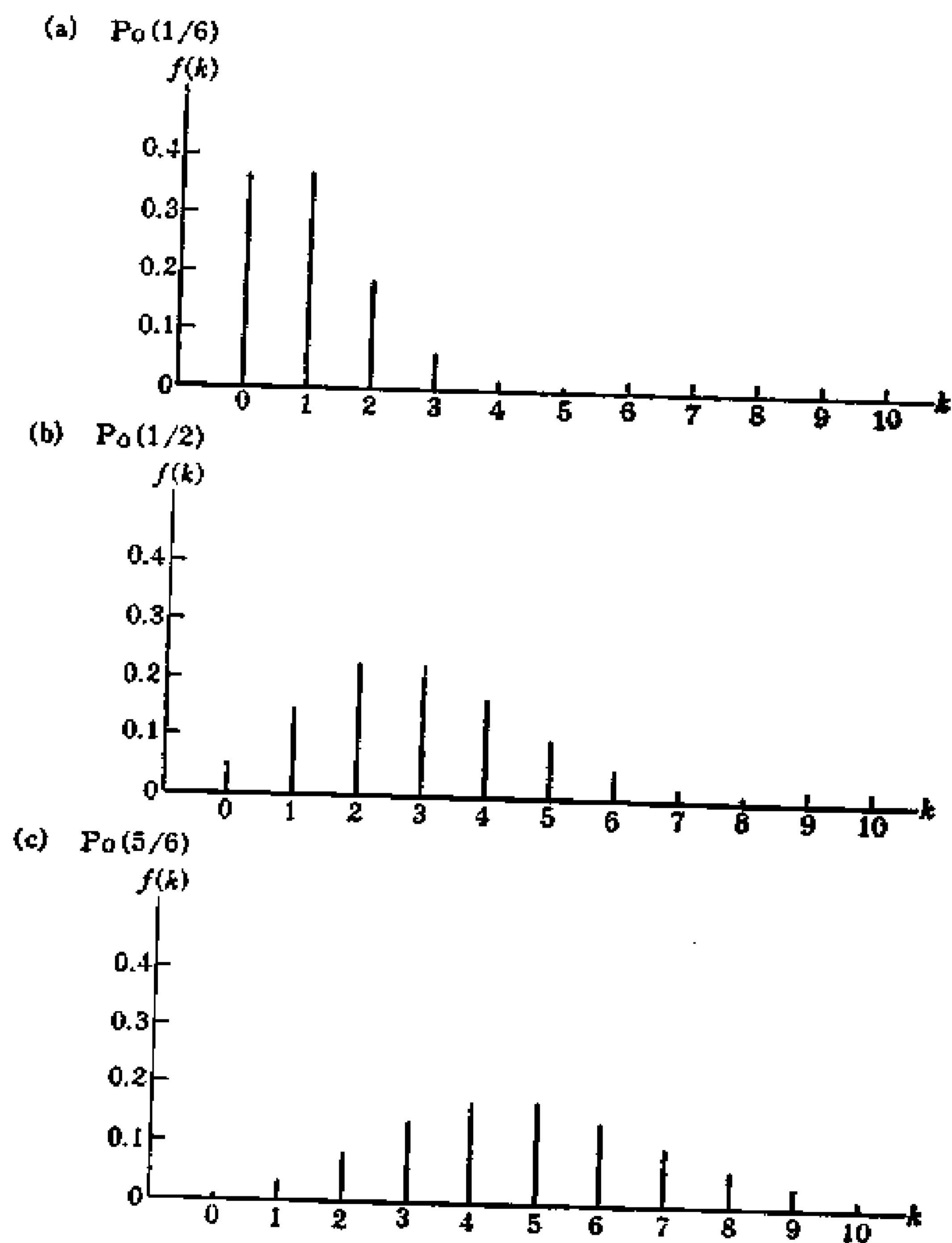


图 2-3 泊松分布的概率分布

成立.

做为例子，表 2-2 给出了将二项分布 B(100; 0.01) 和泊松分布 Po(1) 的概率分布进行对比的结果. 可以看出，当 $n = 100$ 的时候，式 (2.17) 已是一个很好的近似公式.

表 2-2　二项分布的泊松近似

k	$f(k)$	
	B(100;0.01)	Po(1)
0	0.3660	0.3679
1	0.3697	0.3679
2	0.1849	0.1839
3	0.0610	0.0613
4	0.0149	0.0153
5	0.0029	0.0031
6	0.0005	0.0005

2.2.2 连续型分布

在应用中，很多随机变量的取值范围可以是所有实数，所有非负实数或数轴上某一区间内的实数. 其分布函数可写成如下的积分形式

$$F(x) = \int_{-\infty}^{x} f(v)\mathrm{d}v \tag{2.18}$$

这种分布，称做 **连续型概率分布** 或称做 **连续分布** (continuous distribution).

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) \tag{2.19}$$

则称做该分布的 **概率密度函数** (probability density function), 简称 **概率密度**. $f(x)$ 满足以下两个性质 (试将它与离散分布的性质进行比较):

(i) $f(x) \geq 0 \qquad (-\infty < x < \infty).$ (2.20)

(ii) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1.$ (2.21)

具有代表性的连续分布有以下几种：

(1) **均匀分布** (uniform distribution) U(a,b) $(a < b)$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ 0 & (x \text{ 为其它}) \end{cases} \tag{2.22}$$

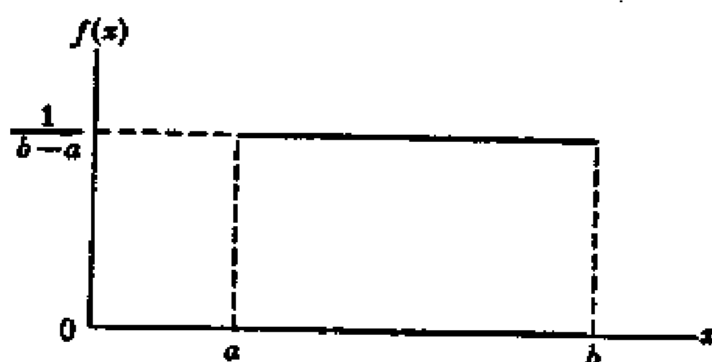


图 2-4 均匀分布的概率密度函数

(2) **贝塔分布** (Beta distribution) Be$(\alpha,\beta)(\alpha > 0, \beta > 0)$

$$f(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & (0 \leq x \leq 1) \\ 0 & (x \text{ 为其它}) \end{cases} \tag{2.23}$$

这里，

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t \tag{2.24}$$

称做**贝塔函数** (Beta function).

当 $\alpha = \beta = 1$ 的时候，贝塔分布变成均匀分布 U(0,1).

当 $\alpha < 1, \beta < 1$ 的时候，概率密度函数在 $0 < x < 1$ 之间呈谷底状 (U 字形).

当 $\alpha > 1, \beta > 1$ 的时候，概率密度函数在 $0 < x < 1$ 之间呈谷峰状.

(3) **指数分布** (exponential distribution) $\mathrm{Ex}(\alpha)$ $(\alpha > 0)$

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \tag{2.25}$$

(4) **伽玛分布** (Gamma distribution) $\mathrm{G}(\alpha, \nu)$ $(\alpha > 0, \quad \nu > 0)$

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\nu)} \alpha^{\nu} x^{\nu-1} e^{-\alpha x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \tag{2.26}$$

这里，

$$\Gamma(\nu) = \int_0^{\infty} t^{\nu-1} e^{-t} \mathrm{d}t \tag{2.27}$$

称做 **伽玛函数** (Gamma function). 伽玛函数具有以下性质

$$\left.\begin{aligned} &\Gamma(\nu) = (\nu - 1)\Gamma(\nu - 1) \\ &\Gamma(1) = 1, \quad \Gamma(\frac{1}{2}) = \sqrt{\pi} \end{aligned}\right\} \tag{2.28}$$

因此，令 n 为自然数，则

$$\left.\begin{aligned} &\Gamma(n) = (n-1)! \\ &\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2})(n - \frac{3}{2})(n - \frac{5}{2}) \cdots \frac{1}{2}\sqrt{\pi} \end{aligned}\right\} \tag{2.29}$$

此外，贝塔函数与伽玛函数之间存在以下关系

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \tag{2.30}$$

当 $\nu = 1$ 的时候，伽玛分布变成指数分布.

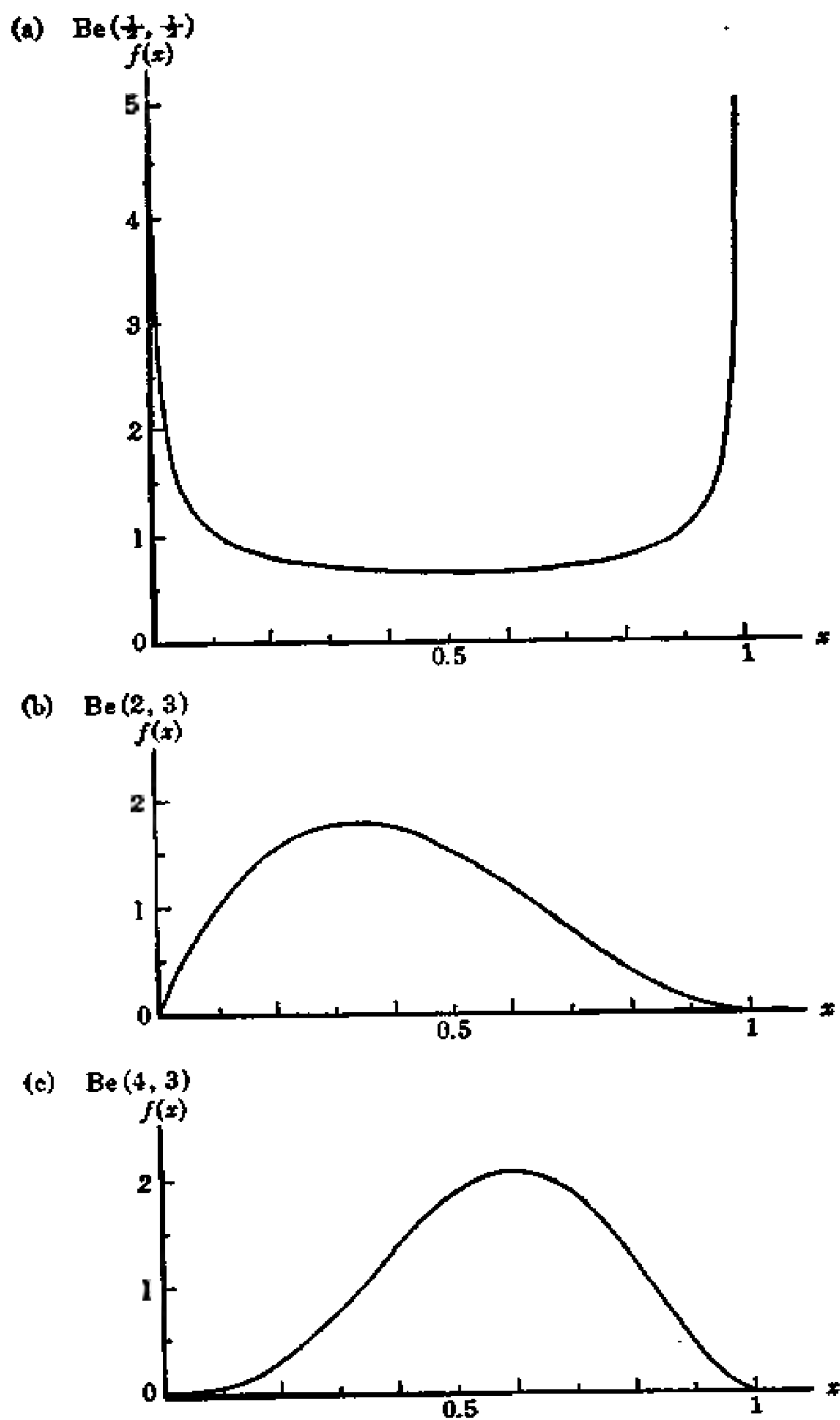


图 2-5 贝塔分布的概率密度函数

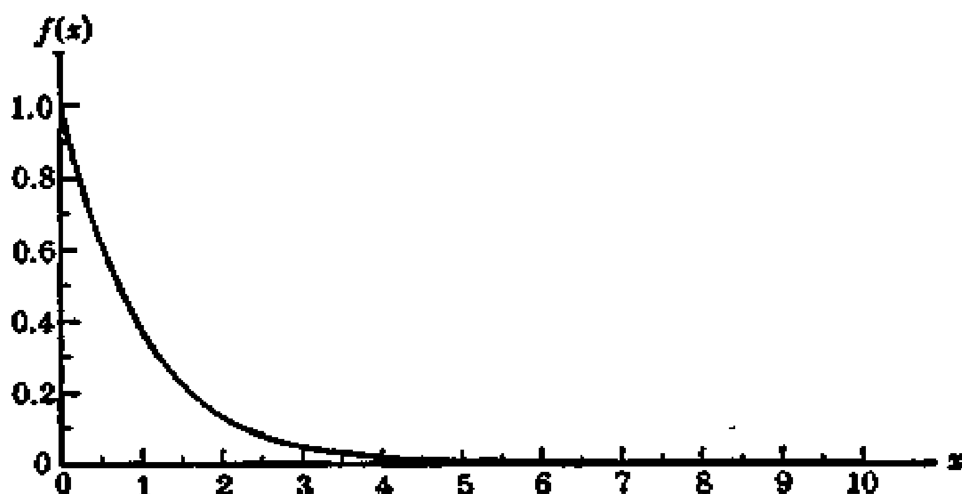


图 2-6 指数分布的概率密度函数

当 α 一定，ν 发生变化的时候，$f(x)$ 的形状会发生变化 (如图 2-7). 因此，我们称 ν 为 **形状参数** (shape paramater). 如果随机变量 X, X' 分别服从伽玛分布 $\mathrm{G}(\alpha,\nu), \mathrm{G}(\alpha',\nu)$ (ν 相同，α 与 α' 不同), 则 αX 与 $\alpha' X'$ 服从同一伽玛分布 $\mathrm{G}(1,\nu)$. 从这个意义上，我们称 α 为伽玛分布的 **尺度参数** (scale parameter).

当 n 为自然数的时候，我们称伽玛分布 $G(1/2, n/2)$ 服从自由度为 n 的 **χ^2 分布** (χ^2 distribution). χ^2 分布在统计学中占有重要地位. 此外，在与排队论有关的理论中，我们有时称伽玛分布服从次数 (phase) 为 n 的 **爱尔朗分布** (Erlang distribution).

(5) **正态分布**(normal distribution) $\mathrm{N}(\mu,\sigma^2)(-\infty < \mu < \infty, \sigma > 0)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \quad (-\infty < x < \infty) \qquad (2.31)^1$$

正态分布在统计学中是最重要的分布之一 (如图 2-8).

特别地，$\mathrm{N}(0,1)$ 称做 **标准正态分布**. 本书中，我们分别用

[1] 当 $g(x)$ 较为复杂的时候，我们用 $\exp[g(x)]$ 表示 $e^{g(x)}$.

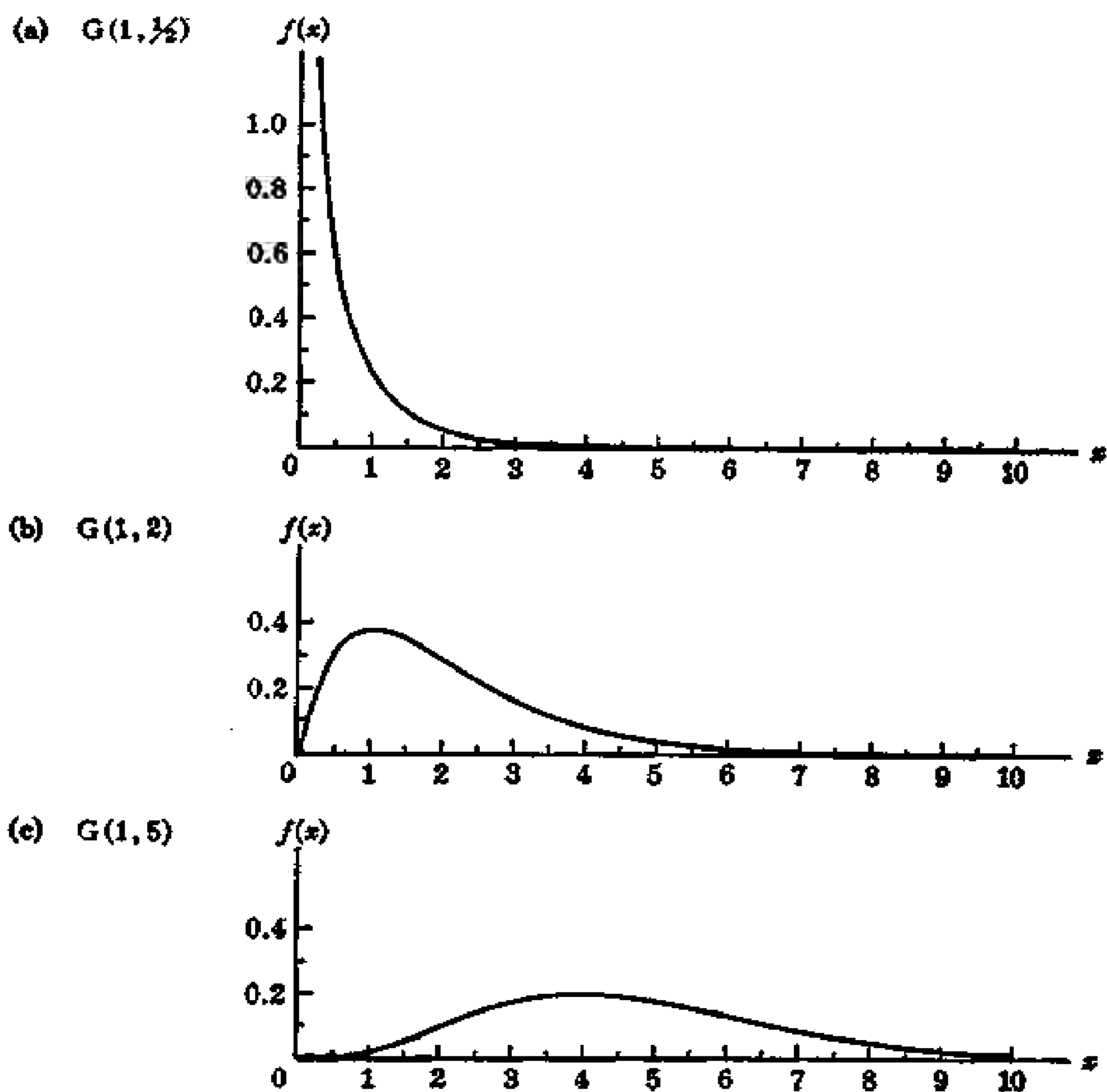


图 2-7　伽玛分布的概率密度函数.

$\phi(x)$ 和 $\Phi(x)$ 来表示其概率密度和分布函数，即

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) \tag{2.32}$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)\mathrm{d}t \tag{2.33}$$

表 2-3 列出了一些与 x 相对应的 $\phi(x)$ 及 $\Phi(x)$ 的值. 详表请参照附录.

图 2-9 给出了 $\phi(x)$ 及 $\Phi(x)$ 的图形. 当 x 取负值的时候，可通过下面的关系式求出 $\phi(x)$ 及 $\Phi(x)$.

$$\phi(x) = \phi(-x) \tag{2.34}$$

$$\Phi(x) = 1 - \Phi(-x) \tag{2.35}$$

问题　设随机变量 X 服从 N(0,1), 试求下面的概率：

(1) $P(|X| \leq 1)$　　(2)$P(|X| \leq 2)$　　(3)$P(|X| \leq 3)$

(4) $P(|X| \leq 1.96)$

一般来讲，服从正态分布 $\mathrm{N}(\mu, \sigma^2)$ 的分布函数 $F(x)$ 的值可由式

$$F(x) = \Phi((x-\mu)/\sigma) \tag{2.36}$$

通过表 2-3(或由附录的详表) 求出. 式 (2.36), 则可由 $F(x)$ 的定义式

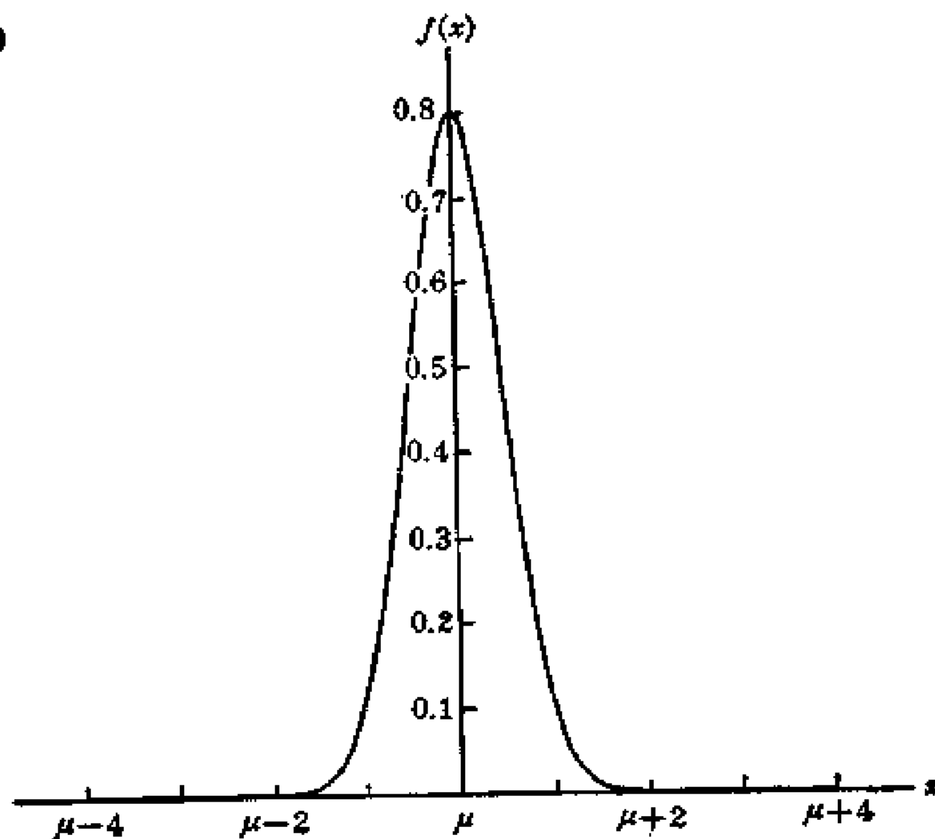
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{y-\mu}{\sigma})^2]\mathrm{d}y$$

通过积分变换 $t = (y-\mu)/\sigma$ 求出.

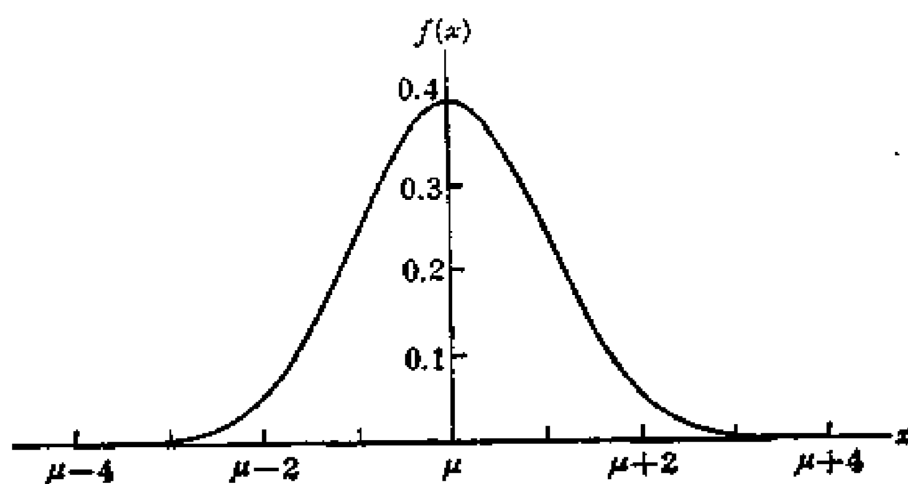
问题　随机变量 Y 服从 N(0,1), 试求以下概率：

(1) $P(-1 \leq Y \leq 4)$　(2) $P(Y \geq 5)$

(a) $N(\mu, 0.5^2)$



(b) $N(\mu, 1^2)$



(c) $N(\mu, 2^2)$

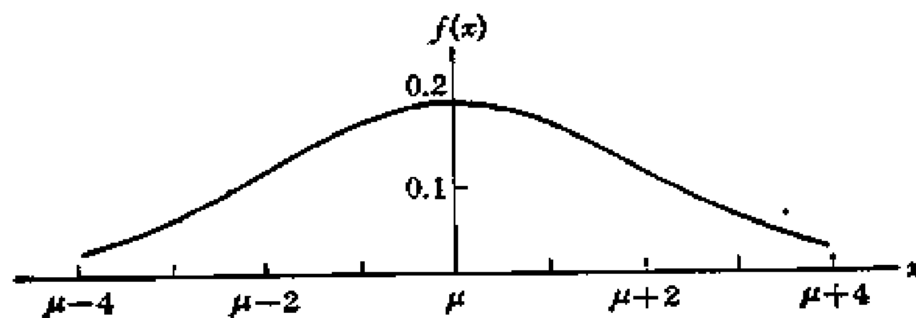


图 2-8　正态分布的概率密度函数

表 2-3 标准正态分布 N(0,1) 的概率密度函数 $\phi(x)$ 与分布函数 $\Phi(x)$ 的值

x	$\phi(x)$	$\Phi(x)$	$1-\Phi(x)$
0	0.399	0.500	0.500
0.5	0.352	0.691	0.309
1.0	0.242	0.841	0.159
1.5	0.130	0.933	0.067
2.0	0.054	0.977	0.023
2.5	0.018	0.994	0.006
3.0	0.004	0.999	0.001

(6) **柯西分布** (Cauchy distribution) $C(\mu,\alpha)$ $(-\infty<\mu<\infty, \alpha>0)$

$$f(x)=\frac{1}{\pi}\cdot\frac{\alpha}{(x-\mu)^2+\alpha^2}\quad(-\infty<x<\infty)\tag{2.37}$$

这个概率密度函数的图形 (图 2-10) 看起来与正态分布很接近，但两者在理论上有很大的性质差异. 这一点，我们将在第 5 章做详细介绍.

正态分布与伽玛分布 (χ^2 分布) 的关系

假定随机变量 X 服从标准正态分布 N(0,1), 我们来求 $Y=X^2$ 的分布函数 $F_Y(y)$ 和概率密度函数 $f_Y(y)$.

当 $y\leq 0$ 的时候，很明显， $F_Y(y)=0$, $f_Y(y)=0$.

当 $y>0$ 的时候，

$$\begin{aligned}F_Y(y)&=P(Y\leq y)=P(X^2\leq y)=P(-\sqrt{y}\leq X\leq\sqrt{y})\\&=\int_{-\sqrt{y}}^{\sqrt{y}}\phi(x)\mathrm{d}x=2\int_0^{\sqrt{y}}\phi(x)\mathrm{d}x\end{aligned}\tag{2.38}$$

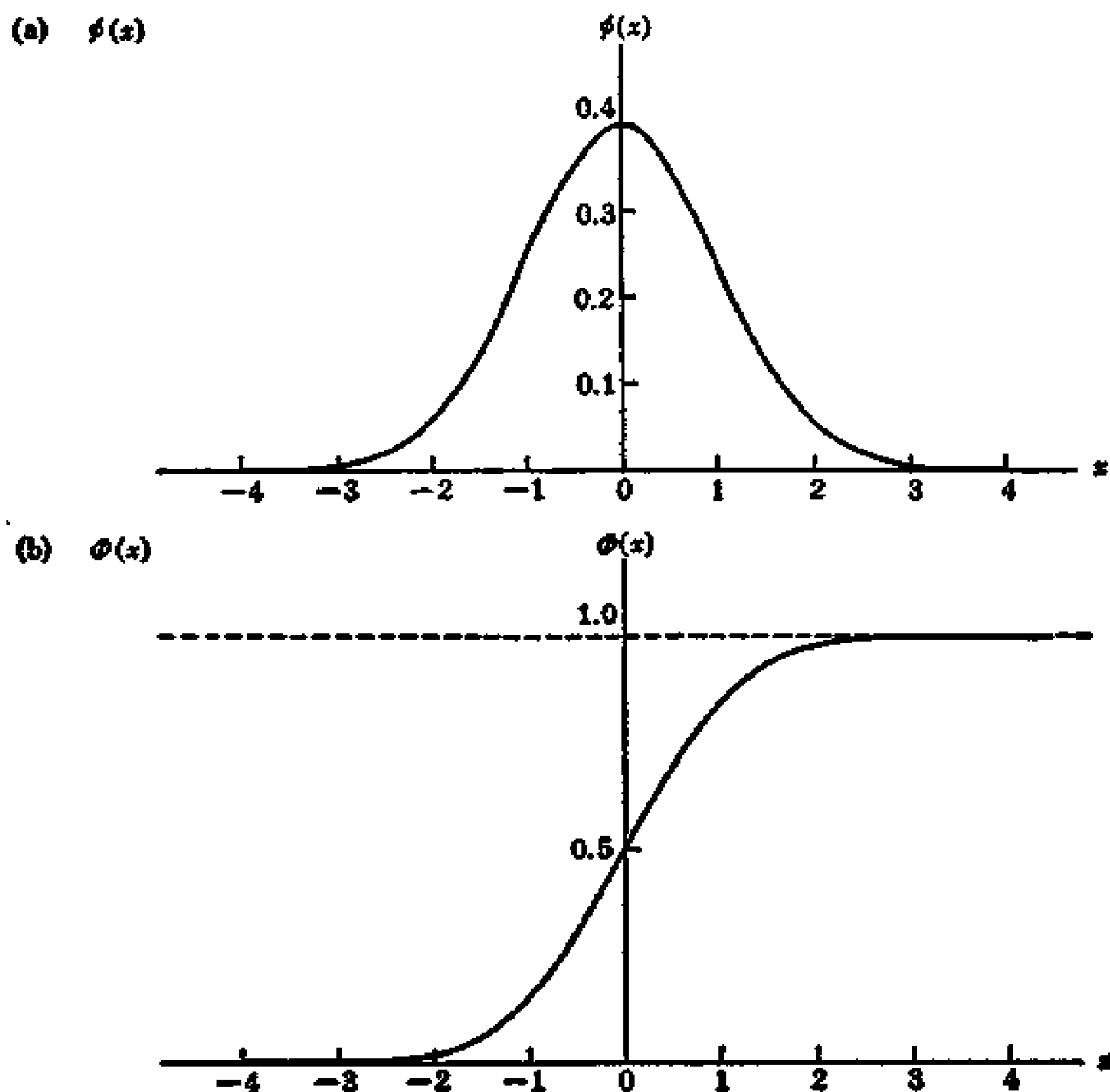


图 2-9 标准正态分布 N(0,1) 的概率密度函数 $\phi(x)$ 与分布函数 $\Phi(x)$

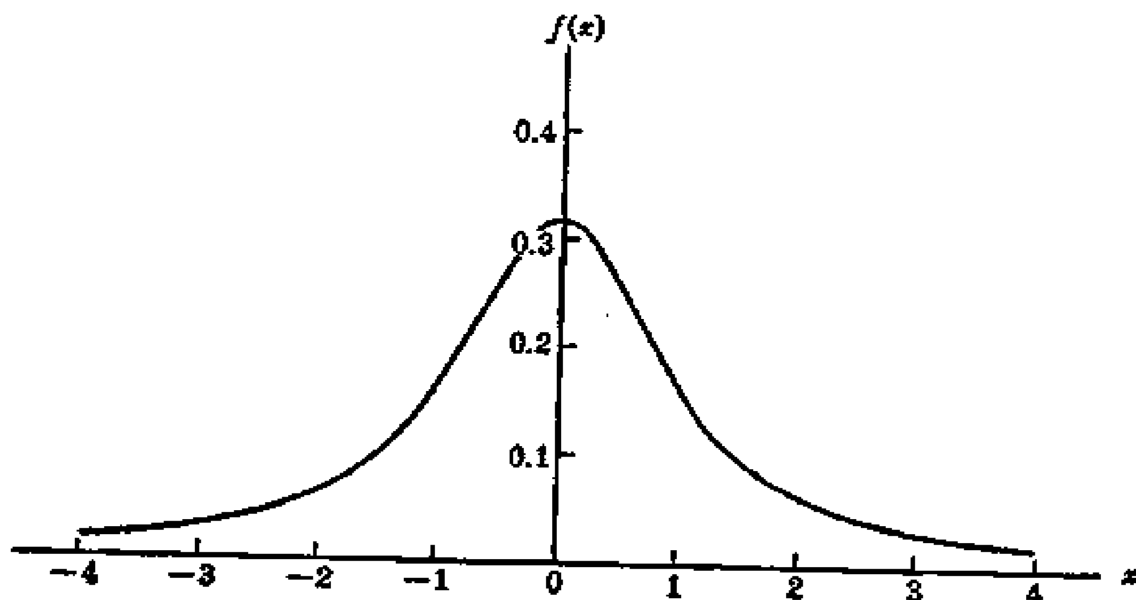


图 2-10 柯西分布 C(0,1) 的概率密度函数

(最后一个等式，由 $\phi(x)$ 为偶函数推知)，因而

$$f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}F_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}\{2\int_0^{\sqrt{y}}\phi(x)\mathrm{d}x\}$$

这里令 $\sqrt{y}=t$，则由复合函数的微分法

$$f_Y(y)=\frac{\mathrm{d}t}{\mathrm{d}y}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\{2\int_0^{t}\phi(x)\mathrm{d}x\}=\frac{1}{2\sqrt{y}}\cdot 2\phi(t)$$

最后用 $\sqrt{y}$ 替换 t 得到

$$f_Y(y)=\frac{1}{\sqrt{2\pi}}y^{-1/2}\exp(-\frac{1}{2}y) \tag{2.39}$$

这是伽玛分布 G(1/2, 1/2)，即它服从自由度为 1 的 χ^2 分布的概率函数.

一般来讲，从服从任意分布的随机变量 X 的概率密度函数求以 X 为自变量的函数的概率密度函数的时候，我们常采用如上以 $\phi(x)$ 推出 $f_Y(y)$ 的思考方法.

问题 设随机变量 X 服从均匀分布 U(0,1), 试证当 α 为正常数的时候, $Y = -\alpha^{-1}\log X$ 服从指数分布 Ex(α).

问题 设随机变量服从连续型分布, 且其概率密度函数及分布函数分别为 $f(x)$ 和 $F(x)$. 试求当 a 为正常数, b 为任意实数的时候, 随机变量 $Y = aX + b$ 的概率密度函数及分布函数.

2.3 随机变量的联合分布

对于某个概率空间 $(\Omega, \mathcal{A}, P)$, 有时我们需要同时考察两个或两个以上的随机变量 $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$, 并研究它们之间的关系. 假如 $x_1, x_2, \cdots, x_n$ 为任意实数值, 则对于所有 $k(1 \leq k \leq n)$ 来讲, $\{\omega : X_k(\omega) \leq x_k\}$ 就是一个事件 (即它属于 $\mathcal{A}$), 因而

$$
\begin{aligned}
&\{\omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \cdots, X_n(\omega) \leq x_n\} \\
&= \bigcap_{k=1}^{n} \{\omega : X_k \leq x_k\}
\end{aligned}
$$

也属于 $\mathcal{A}$. 这样, 我们就可以算出它们对应的概率. 这个概率, 可写成 $F_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n)$ 或简写成 $F(x_1, x_2, \cdots, x_n)$ 的形式, 并称它为随机变量 $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ 的 **联合分布函数** (joint distribution function) 或简称 **分布函数**. 即

$$
F(x_1, x_2, \cdots, x_n) = P(\bigcap_{k=1}^{n} \{\omega : X_k(\omega) \leq x_k\}) \tag{2.40}
$$

下面具体讨论一下 $n = 2$ 的时候, 联合分布函数的性质.

我们把每个随机变量 $X_1(\omega), X_2(\omega)$ 看做是在直线上随机走动的质点. 同理, 我们把二维随机变量 $(X_1(\omega), X_2(\omega))$ 看做是在某个平面上随机走动的质点 (如图 2-11 所示).

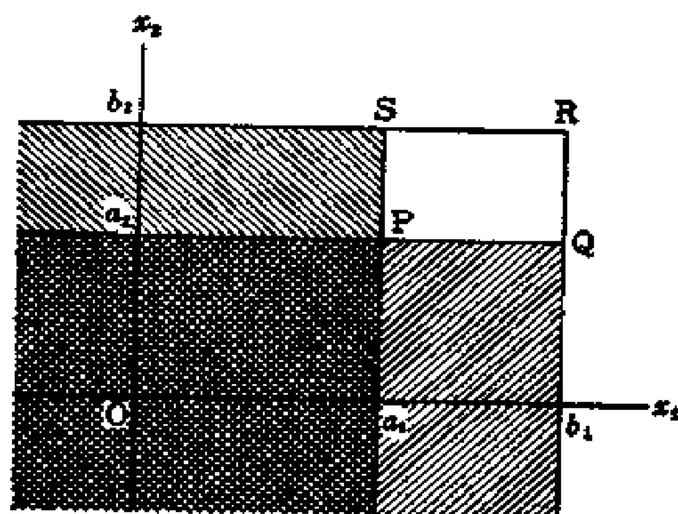


图 2-11　二元随机变量的联合分布

即如果 $X_1(\omega)$ 的取值为 a_1, $X_2(\omega)$ 的取值为 a_2, 这个质点可看做是位于坐标 (a_1,a_2) 的点 P.

这样，按照上面的解释，对应于这一点的联合分布函数的值 $F(a_1,a_2)$ (为了方便，我们把 $F(a_1,a_2)$ 简记为 $F(P)$), 就是这个质点位于图 2-11 所示右上斜线和右下斜线重叠区域

$$\{(x_1,x_2): -\infty < x_1 \le a_1, \quad -\infty < x_2 \le a_2\}$$

的概率. 同理，质点位于右上斜线区域

$$\{(x_1,x_2): a_1 < x_1 \le b_1, \quad -\infty < x_2 \le a_2\}$$

的概率为

$$F(\mathrm{Q}) - F(\mathrm{P}) = F(b_1,a_2) - F(a_1,a_2)$$

质点位于右下斜线区域

$$\{(x_1,x_2): -\infty < x_1 \le a_1, \quad a_2 < x_2 \le b_2\}$$

的概率为

$$F(\mathrm{S}) - F(\mathrm{P}) = F(a_1, b_2) - F(a_1, a_2)$$

特别地，质点位于矩形区域 PQRS

$$\{(x_1, x_2) : a_1 < x_1 \le b_1, \quad a_2 < x_2 \le b_2\}$$

的概率

$$\begin{aligned} &F(\mathrm{R}) - F(\mathrm{P}) - \{F(\mathrm{Q}) - F(\mathrm{P})\} - \{F(\mathrm{S}) - F(\mathrm{P})\} \\ &= F(\mathrm{R}) - F(\mathrm{Q}) - F(\mathrm{S}) + F(\mathrm{P}) \end{aligned}$$

由此，我们可以推出下面所述分布函数的性质 (i). 性质 (ii) 及性质 (iii) 可仿一维随机变量情况推出.

(i) 单调不减性：如果 $a_1 \le b_1$, $a_2 \le b_2$, 则

$$F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \ge 0 \tag{2.41}$$

(ii) 右连续性：

$$\lim_{\substack{x_1 \to a_1 + 0 \\ x_2 \to a_2 + 0}} F(x_1, x_2) = F(a_1, a_2) \tag{2.42}$$

(iii)

$$\lim_{x_1 \to -\infty} F(x_1, x_2) = \lim_{x_2 \to -\infty} F(x_1, x_2) = 0 \tag{2.43}$$

$$\lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1, x_2) = 1 \tag{2.44}$$

$F_1(x) \equiv F(x, +\infty)$ 是一维随机变量的分布函数 ($X_2(\omega)$ 的取值不去考虑), 它表示 $X_1(\omega)$ 小于等于 x 的概率. 我们称 $F_1(x)$ 为由 $F(x_1, x_2)$ 确定的，关于 $X_1(\omega)$ 的 **边缘分布函数** (marginal distribution function). 同理， $F_2(x) \equiv F(+\infty, x)$ 是关于 $X_2(\omega)$ 的边缘分布函数.

此外对于固定的 c, 我们称 $F(x,c)/F(\infty,c)$ 为条件 $X_2(\omega)\le c$ 下的关于 $X_1(\omega)$ 的 **条件分布函数** (conditional distribution).

例 2-2 二维正态分布(bivariate normal contribution) 如果二维随机变量的联合分布可以写成

$$F(x_1,x_2)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}f(u,v)\mathrm{d}v\mathrm{d}u \tag{2.45}$$

的形式, 我们称它服从连续分布, 并称 $f(u,v)$ 为其概率密度函数. 其中, 特别常用的是二维正态分布

$$f(u,v)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp[-\frac{1}{2}Q(u,v)] \qquad ((2.46))$$

$$Q(u,v)=\frac{1}{1-\rho^2}\{(\frac{u-\mu_1}{\sigma_1})^2 -2\rho(\frac{u-\mu_1}{\sigma_1})(\frac{v-\mu_2}{\sigma_2})+(\frac{v-\mu_2}{\sigma_2})^2\}$$

这里, σ_1,σ_2 为正常数, ρ 为满足 $-1<\rho<1$ 的常数. 如果假定 $f_1(x)$ 是关于 $X_1(\omega)$ 的边缘分布的概率密度函数, 则

$$f_1(x)=\frac{\mathrm{d}}{\mathrm{d}x}F_1(x) \tag{2.47}$$

$$F_1(x)=\int_{-\infty}^{x}\{\int_{-\infty}^{\infty}f(u,v)\mathrm{d}v\}\mathrm{d}u \tag{2.48}$$

故

$$f_1(x)=\int_{-\infty}^{\infty}f(x,v)\mathrm{d}v \tag{2.49}$$

把上式代入式 (2.46), 积分 (比较繁琐) 得到

$$f_1(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp[-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2] \tag{2.50}$$

很显然, $X_1(\omega)$ 的边缘分布为正态分布 $\mathrm{N}(\mu,\sigma_1^2)$. 同理, $X_2(\omega)$ 的边缘分布也为正态分布 $\mathrm{N}(\mu_2,\sigma_2^2)$.

当考察两个有关系的变量之间的分布，比如学生(身长，椅高)分布，(数学成绩，英语成绩)的分布的时候，我们常会用到二维正态分布.

2.4 随机变量的独立性

如果 n 维随机变量 $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ 的联合分布函数 $F(x_1, x_2, \cdots, x_n)$ 可以写成 n 个一维随机变量分布函数的积的形式

$$F(x_1, x_2, \cdots, x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n) \tag{2.51}$$

我们称这 n 个随机变量是 **相互独立** 的. 这里，每个 $F_k(x_k)$ 称做此分布的边缘分布函数. 当 n 个随机变量均服从离散或连续分布的时候，我们可用概率分布或概率密度函数来代替分布函数，并把式 (2.51) 改写成

$$f(x_1, x_2, \cdots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n) \tag{2.52}$$

这里，$f_k(x_k)$ 是边缘分布 $F_k(x_k)$ 的概率分布或概率密度函数.

例 2-3 在二维正态分布的式 (2.46) 中，如果令 $\rho = 0$, 则

$$\begin{aligned}
f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2}\exp[-\frac{1}{2}\{(\frac{x_1-\mu_1}{\sigma_1})^2 + (\frac{x_2-\mu_2}{\sigma_2})^2\}] \\
&= f_1(x_1)f_2(x_2) \\
f_i(x_i) &= \frac{1}{\sqrt{2\pi}\sigma_i}\exp[-\frac{1}{2}(\frac{x_i-\mu_i}{\sigma_i})^2] \quad (i = 1, 2)
\end{aligned}$$

很显然，$X_1(\omega)$ 和 $X_2(\omega)$ 相互独立且分别服从正态分布.

例 2-4 X_1 服从二项分布 $\mathrm{B}(n_1; p)$, X_2 服从二项分布 $\mathrm{B}(n_2; p)$, 且它们相互独立. 试证 $S = X_1 + X_2$ 也服从二项分布 $\mathrm{B}(n_1 + n_2; p)$.

根据式 (2.51) 及全概率公式，对于所有满足 $k_1+k_2=k$ 的组合 (k_1,k_2)，我们只要将

$$\begin{pmatrix} n_1 \\ k_1 \end{pmatrix} p^{k_1}(1-p)^{n_1-k_1} \times \begin{pmatrix} n_2 \\ k_2 \end{pmatrix} p^{k_2}(1-p)^{n_2-k_2} \tag{2.53}$$

进行累加，即可得到 $S=k$ 的概率．由于式 (2.53) 相当于

$$\begin{pmatrix} n_1 \\ k_1 \end{pmatrix}\begin{pmatrix} n_2 \\ k_2 \end{pmatrix} p^k(1-p)^{n-k} \qquad (n=n_1+n_2)$$

又因为

$$\sum_{k_1+k_2=k}\begin{pmatrix} n_1 \\ k_1 \end{pmatrix}\begin{pmatrix} n_2 \\ k_2 \end{pmatrix} = \begin{pmatrix} n \\ k \end{pmatrix} \tag{2.54}$$

成立 [1]，所以

$$P(S=k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \tag{2.55}$$

即， S 服从 $\mathrm{B}(n;P)$.

问题 设随机变量 X_1, X_2 分别服从泊松分布 $\mathrm{Po}(\lambda_1)$, $\mathrm{Po}(\lambda_2)$，且它们相互独立．试证 X_1+X_2 也服从泊松分布 $\mathrm{Po}(\lambda_1+\lambda_2)$.

提示：可考虑利用二项展式

$$(\lambda_1+\lambda_2)^k = \sum_{k_1=0}^{k}\begin{pmatrix} k \\ k_1 \end{pmatrix}\lambda_1^{k}\lambda_2^{k-k_1} \tag{2.56}$$

---

[1] 把 n 个物体分成 n_1 和 $n_2(=n-n_1)$ 两个部分．当 $k_1=0,1,2,\cdots,k$ 的时候，把从一部分取出 k_1 个物体，另一部分取出 k_2 个物体的组合数逐次进行累加，就会得到从 n 个物体中取出 k 个物体的组合数．

2.5 相互独立随机变量的和的分布

上一节中，我们求出了服从二项分布及泊松分布的两个相互独立的随机变量的和的分布．这一节里，我们把它推广到更一般的情况，考察服从任意分布的、相互独立随机变量的和的分布．

设随机变量 X_1, X_2 相互独立且服从离散分布．它们的概率分布分别为 $f_1(\cdot)$ 和 $f_2(\cdot)$．令 X_1 所取的值的集合 $V = \{v_0, v_1, \cdots\}$．如假定 $S = X_1 + X_2$ 的概率分布为 $f(\cdot)$, 则仿上一节例 2-4,

$$\begin{aligned} f(x) = P(X_1 + X_2 = x) &= \sum_{v \in V} P(X_1 = v)P(X_2 = x - v) \\ &= \sum_{v \in V} f_1(v)f_2(x - v) \end{aligned} \tag{2.57}$$

式 (2.57) 右边的积和形式 (它的特点在于无论 x 的取值如何，f_1 的自变量与 f_2 的自变量之和将保持不变), 称做 f_1 与 f_2 的卷积 (convolution). 特别地，如果随机变量 X_1, X_2 只取非负整数值，则式 (2.57) 将变为

$$f(x) = \sum_{k=0}^{x} f_1(k)f_2(x - k) = \sum_{k=0}^{x} f_1(x - k)f_2(x) \tag{2.58}$$

下面再讨论连续型随机变量的和的分布．设随机变量 X_1, X_2 相互独立且服从连续型分布．它们的概率密度函数分别为 $f_1(\cdot)$ 和 $f_2(\cdot)$, 分布函数分别为 $F_1(\cdot)$ 和 $F_2(\cdot)$. 如假定 $S = X_1 + X_2$ 的概率密度函数及分布函数分别为 $f(\cdot)$ 和 $F(\cdot)$, 则

$$\begin{aligned} F(x) = P(X_1 + X_2 \le x) &= \int\int_{x_1 + x_2 \le x} f_1(x)f_2(x)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty}\{\int_{-\infty}^{x - x_1} f_2(x)\mathrm{d}x_2\}f_1(x_1)\mathrm{d}x_1 \\ &= \int_{-\infty}^{\infty} F_2(x - x_1)f_1(x_1)\mathrm{d}x_1 \end{aligned} \tag{2.59}$$

从上式第一行推到第二行的过程中，先对 x_2 求积，然后再对 x_1 求积. 下面变换一下求积的顺序，则

$$F(x) = \int_{-\infty}^{\infty} F_1(x - x_2) f_2(x_2) \mathrm{d}x_2 \tag{2.60}$$

对式 (2.59) 及式 (2.60) 关于 x 求导，我们得到

$$f(x) = \int_{-\infty}^{\infty} f_2(x - x_1) f_1(x_1) \mathrm{d}x_1 = \int_{-\infty}^{\infty} f_1(x - x_2) f_2(x_2) \mathrm{d}x_2 \tag{2.61}$$

即对于连续型分布来讲，和的分布的概率密度函数可通过各个分布的概率密度函数的 (积分形式的) 卷积求出.

例 2-5 设相互独立的随机变量 X_1, X_2 服从同一指数分布 $\mathrm{Ex}(\alpha)$. 试求它们的和的分布.

因为 $f_1(t) = f_2(t) = \alpha e^{-\alpha t} \quad (t \geq 0)$; $f_1(t) = f_2(t) = 0.\ (t < 0)$, 故若 $x < 0$, 则 $f(x) = 0$, 若 $x > 0$, 则

$$f(x) = \int_0^x \alpha e^{-\alpha(x-t)} \cdot \alpha e^{-\alpha t} \mathrm{d}t = \alpha^2 e^{-\alpha x} \int_0^x \mathrm{d}t = \alpha^2 x e^{-\alpha x}$$

这是伽玛分布 $\mathrm{G}(\alpha, 2)$ 的概率密度函数.

例 2-6 设相互独立的随机变量 X_1, X_2 分别服从伽玛分布 $\mathrm{G}(\alpha, \nu_1)$ 和 $\mathrm{G}(\alpha, \nu_2)$. 试证 $X_1 + X_2$ 也服从伽玛分布 $\mathrm{G}(\alpha, \nu_1 + \nu_2)$.

当 $x > 0$ 的时候，

$$\begin{aligned} f(x) &= \int_0^x \frac{\alpha^{\nu_1}}{\Gamma(\nu_1)} (x - t)^{\nu_1 - 1} e^{-\alpha(x-t)} \cdot \frac{\alpha^{\nu_2}}{\Gamma(\nu_2)} t^{\nu_2 - 1} e^{-\alpha t} \mathrm{d}t \\ &= \frac{\alpha^{\nu_1 + \nu_2} e^{-\alpha x}}{\Gamma(\nu_1)\Gamma(\nu_2)} \int_0^x (x - t)^{\nu_1 - 1} t^{\nu_2 - 1} \mathrm{d}t \end{aligned}$$

进行积分变换，即从 t 变到 $s=t/x$, 则

$$\begin{aligned}\int_0^x (x-t)^{\nu_1-1}t^{\nu_2-1}\mathrm{d}t &= \int_0^1 (x-sx)^{\nu_1-1}(sx)^{\nu_2-1}x ds \\ &= x^{\nu_1+\nu_2-1}\int_0^1 (1-s)^{\nu_1-1}s^{\nu_2-1}ds = x^{\nu_1+\nu_2-1}B(\nu_1,\nu_2) \\ &= \frac{\Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_1+\nu_2)}x^{\nu_1+\nu_2-1}\end{aligned}$$

最后求出

$$f(x)=\frac{\alpha^{\nu_1+\nu_2}}{\Gamma(\nu_1+\nu_2)}x^{\nu_1+\nu_2-1}e^{-\alpha x}\quad (x>0)$$

这是伽玛分布 $G(\alpha,\nu_1+\nu_2)$ 的概率密度函数.

按同样的方法，我们可以求出三个及三个以上相互独立随机变量的和的分布. 为了方便，我们简单地用

$$f=f_1*f_2$$

表示离散分布公式 (2.57) 及连续分布公式 (2.58) 的卷积. 根据前面所述，我们已知卷积演算 * 满足交换律

$$f_1*f_2=f_2*f_1 \tag{2.62}$$

现设三个随机变量 X_1 ,X_2, X_3 的概率分布或概率密度函数分别为 f_1, f_2, f_3, $S=X_1+X_2+X_3$ 的概率分布或概率密度函数为 f. 则对于 $S=(X_1+X_2)+X_3$ 有

$$f=(f_1*f_2)*f_3$$

又对于 $S=X_1+(X_2+X_3)$ 有

$$f=f_1*(f_2*f_3)$$

可见，卷积演算 * 还满足结合律

$$(f_1 * f_2) * f_3 = f_1 * (f_2 * f_3) \tag{2.63}$$

即我们可以舍去括号而把它简写成

$$f = f_1 * f_2 * f_3$$

一般地，对于 n 个随机变量，我们同样有

$$f = f_1 * f_2 * f_3 * \cdots * f_n \tag{2.64}$$

特别地，如果 n 个随机变量服从同一分布，即如果 $f_1 = f_2 = \cdots = f_n$，则和的分布可写成

$$f = \underbrace{f_1 * f_1 * \cdots * f_1}_{f_1 \text{有} n \text{个}} \equiv f_1^{(n)} \tag{2.65}$$

的形式. 上式的右边称做 f_1 的 n 重卷积，并把它表示成为 $f_1^{(n)}$. 此外，如果 f_1 对应的分布函数为 F_1，则可将 $f_1^{(n)}$ 对应的分布函数写成 $F_1^{(n)}$，并称 $F_1^{(n)}$ 为 F_1 的 n 重卷积 (见第 6 章). 即如果相互独立的随机变量 $X_1, X_2, \cdots, X_n$ 服从同一分布 F_1，则

$$P(X_1 + X_2 + \cdots + X_n \leq x) = F_1^{(n)}(x) \tag{2.66}$$

习 题 2

1. 设 X, Y 是相互独立的随机变量，它们分别服从参数为 λ_1, λ_2 的泊松分布，试证

$$P(X = k | X + Y = l) = \begin{pmatrix} l \\ k \end{pmatrix} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{l-k}$$

$$(k = 0, 1, \cdots, l)$$

2. 设随机变量 Z 服从参数为 λ 的泊松分布 $\mathrm{Po}(\lambda)$, 随机变量 X 在 $Z=l$(l 为任意非负整数) 的条件下服从二项分布 $\mathrm{B}(l;p)$. 试求 X 的分布.

3. 设随机变量 X 服从几何分布 $\mathrm{Ge}(p)$. Y 服从指数分布 $\mathrm{Ex}(\alpha)$.

(1) 假定 l 为任意自然数, 试求 $X \geq l$ 条件下 X 的条件概率分布.

(2) 假定 Z 为任意正数, 试求 $Y \geq Z$ 条件下 Y 的条件概率分布.

4. 设随机变量 X 只取非负整数值. 若对任意非负整数 m,n 有 $P(X \geq n+m|X \geq n) = P(X \geq m)$, 试证 X 服从几何分布.

5. 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立且服从均匀分布 $\mathrm{U}(0,1)$, 试求 $Y = -\log(X_1 X_2 \cdots X_n)$ 的分布.

6. 设 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量, 它们的分布函数是 F. 试求 $Y = \max(X_1, X_2, \cdots, X_n)$, $Z = \min(X_1, X_2, \cdots, X_n)$ 的分布函数 (这是信赖性理论领域中的重要课题. 设构成系统的同一种类的 n 个零件的寿命分别为 $X_1, X_2, \cdots, X_n$, 如果系统是串联的, 则其寿命为 Y; 如果系统是并联的, 则其寿命为 Z).

7. 设连续型随机变量 X 的分布函数为 F. 试证随机变量 $Y = F(X)$ 服从均匀分布 $\mathrm{U}(0,1)$ (无论 F 为何分布函数).

8. 设只取正值的连续型随机变量 X, Y 的联合分布的概率密度函数为 $f(x,y)$. 试求 $U = X+Y$, $V = X-Y$, $W = XY$ 及 $Z = Y/X$ 的概率密度函数.

9. 二维正态分布的概率密度函数由式 (2.46) 给出. 试求其边缘概率密度函数.

10. 设随机变量 X_1, X_2, X_3 相互独立且服从正态分布 $\mathrm{N}(0,\sigma^2)$. 试求 $V = \sqrt{X_1^2 + X_2^2 + X_3^2}$ 的概率密度函数 (V 的分布被称为**马克斯威尔分布** (Maxwell distribution), 在物理学中常用于做气体分

子运动论的模型.)

11. 设 r 为自然数，p 为 0 到 1 之间的常数. $q=1-p$, 则

$$f(k)=\binom{r+k-1}{k}p^r q^k \quad (k=0,1,2,\cdots)$$

称做**负二项分布** (negative binomial distribution) $\mathrm{NB}(r;p)$ 的概率分布.

(1) 试证在成功的概率为 p, 失败的概率为 q 的多重伯努利试验中，直到首次成功 r 次所需失败试验次数为 k 次的概率为 $f(k)$.

(2) 二项系数

$$\binom{n}{k}=\frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 1}$$

本来用于表示组合数，因而只适用于满足 $0\le k\le n$ 的整数 n,k. 假定对于任意 $n,k(k\ge 0)$ 均可用上面的定义，试证以下等式成立 (分布的名称由此而来).

$$\binom{r+k-1}{k}=(-1)^k\binom{-r}{k}, \quad f(k)=\binom{-r}{k}p^r(-q)^k$$

12. 在信赖性理论中经常用到的**威布尔分布** (Weibull distribution) 的概率密度为

$$f(x)=\alpha m x^{m-1}\exp[-\alpha x^m] \qquad (x\ge 0,\alpha>0,m>0)$$

试问如果 X 服从此分布，则 $Y=X^m$ 服从什么分布.

第 3 章　随机变量的数字特征

3.1　数学期望

3.1.1　离散型

设离散型随机变量 X 的概率分布为

$$f(v_k) = P(X = v_k) \qquad (k = 0, 1, 2, \cdots).$$

这时，我们称

$$E(X) \equiv \sum_k v_k f(v_k) \tag{3.1}$$

为此随机变量或此分布的 **均值** 或 **数学期望** (expectation).

注　记号 E 来自数学期望的英译词 expectation 头字母. 严格地讲，当式 (3.1) 的级数绝对收敛的时候，即当且仅当条件

$$\sum_k |v_k| f(v_k) < \infty \tag{3.2}$$

成立的时候，我们将式 (3.1) 称做均值，而在其它情况下则认为均值不存在. 由于实际出现的绝大多数分布均能满足此条件，我们一般不去考虑它.

下面求曾在第 2 章第 2 节出现过的几个离散分布的均值.

(1) 0 - 1 分布 B(1; p)

$$E(X) = 0 \times f(0) + 1 \times f(1) = p$$

(2) 二项分布 B$(n;p)$

$$E(X)=\sum_{k=0}^{n}kf(k)=\sum_{k=1}^{n}kf(k)=\sum_{k=1}^{n}k\begin{pmatrix}n\\k\end{pmatrix}p^k(1-p)^{n-k}$$

当 $k\geq 1$ 的时候，

$$k\begin{pmatrix}n\\k\end{pmatrix}=k\frac{n!}{k!(n-k)!}=\frac{n!}{(k-1)!(n-k)!}=n\frac{(n-1)!}{(k-1)!(n-k)!}$$

$$=n\frac{m!}{l!(m-l)!}=n\begin{pmatrix}m\\l\end{pmatrix}$$

这里， $l=k-1, m=n-1$. 因而

$$E(X)=\sum_{l=0}^{m}n\begin{pmatrix}m\\l\end{pmatrix}p^{l+1}(1-p)^{m-l}$$

$$=np\sum_{l=0}^{m}\begin{pmatrix}m\\l\end{pmatrix}p^l(1-p)^{m-l}=np$$

因为 $\begin{pmatrix}m\\l\end{pmatrix}p^l(1-p)^{m-l}$ 为二项分布 B$(m;p)$ 的概率分布，所以把此式从 $l=0$ 加到 $l=m$ 的时候其和为 1. 由此我们可以推出上面的最后一个等式.

(3) 几何分布 Ge(p)

$$E(X)=\sum_{k=0}^{\infty}k(1-p)^kp=\frac{(1-p)}{p}$$

这里，用到了级数公式

$$\sum_{k=0}^{\infty}kx^k=\frac{x}{(1-x)^2}\quad(|x|<1)\tag{3.3}$$

(4) 泊松分布 Po(λ)

$$E(X) = \sum_{k=0}^{\infty} kf(k) = \sum_{k=1}^{\infty} kf(k) = \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!}$$
$$= \lambda\sum_{l=0}^{\infty} e^{-\lambda}\frac{\lambda^l}{l!} = \lambda\sum_{l=0}^{\infty} f(l) = \lambda$$

这里, $l = k - 1$.

设 $g(x)$ 为任意实函数, 则 $g(X)$ 亦为随机变量. 假定 $g(X)$ 的取值为 $u_0, u_1, u_2, \cdots$, 概率分布为 $h(u)$, 则由上面的定义, $g(x)$ 的数学期望为

$$E(g(X)) = \sum_j u_j h(u_j) \tag{3.4}$$

利用下面的定理, 我们不用求出 $g(X)$ 的概率分布而可以通过随机变量 X 的概率分布直接算出 $g(X)$ 的数学期望.

定理 3-1 [1] $E(g(X)) = \sum_k g(v_k)f(v_k)$ (3.5)

当 X 的取值 v_k 皆为非负的时候, 也可由分布函数 F 代替概率分布 f 算出 $E(X)$.

定理 3-2 如果随机变量 X 只取非负值, 则

$$E(X) = \int_0^{\infty} P(X > x)\mathrm{d}x = \int_0^{\infty} \{1 - F(x)\}\mathrm{d}x \tag{3.6}$$

证明 $E(X) = \sum_k v_k f(v_k)$ 相当于图 3-1 中阴影部分的面积. 这个面积又等于 $\int_0^{\infty}\{1 - F(x)\}\mathrm{d}x$.

---

[1] 此定理的证明, 参考习题 3 问题 5 及其解答.

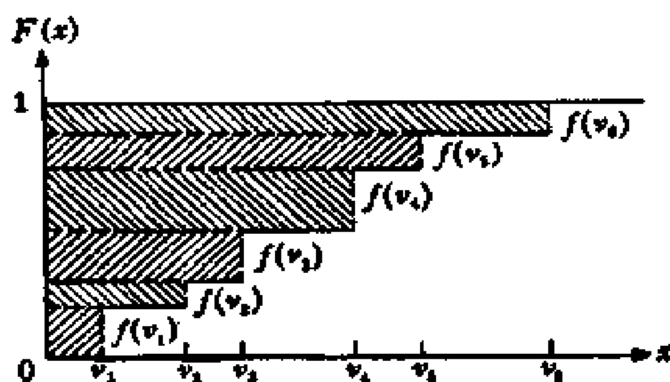


图 3-1 只取非负值的随机变量的数学期望

推论 如果随机变量 X 只取非负整数值，则

$$E(X) = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty}\{1 - F(k)\} = \sum_{k=1}^{\infty} P(X \geq k) \tag{3.7}$$

如果随机变量 X 只取负值，则

$$E(X) = \sum_{k} v_k f(v_k) = -\sum_{k} |v_k| f(v_k)$$

因为 $\sum\limits_{k} |v_k| f(v_k)$ 相当于图 3-2 中阴影部分的面积，所以

$$E(X) = -\int_{-\infty}^{0} P(X \leq x)\mathrm{d}x = -\int_{-\infty}^{0} F(x)\mathrm{d}x \tag{3.8}$$

一般来讲，当随机变量既可取正值又可取负值的时候，有如下定理.

定理 3-3

$$E(X) = \int_{0}^{\infty} P(X > x)\mathrm{d}x - \int_{-\infty}^{0} P(X \leq x)\mathrm{d}x \tag{3.9}$$

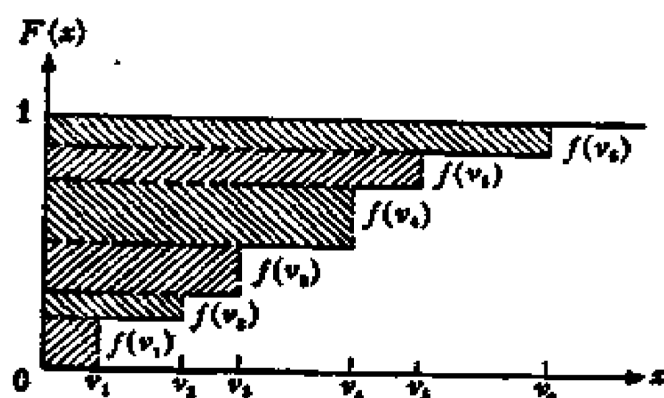


图 3-2 只取负值的随机变量的数学期望

条件期望 (conditional expectation) 设对于离散型随机变量 X 与 Y, $P(Y=y)>0$, 则在事件 $Y=y$ 发生的条件下 X 服从概率分布

$$f(v|y)=P(X=v|Y=y)=\frac{P(X=v,Y=y)}{P(Y=y)} \tag{3.10}$$

另外，在 $Y=y$ 条件下 X 的数学期望

$$E(X|Y=y)=\sum_k v_k f(v_k|y) \tag{3.11}$$

于是，在式 (3.10) 的两边乘以 $P(Y=y)$, 并对所有可能的 Y 值进行累加，便可得到 X 的 (无条件) 数学期望．即

定理 3-4 $$E(X)=\sum_y E(X|Y=y)P(Y=y) \tag{3.12}$$

此式可由全概率公式推出 (第 1 章式 (1.21))．由于简便易用，我们在以后的章节中经常用到这个定理．

3.1.2 连续型

设连续型随机变量 X 的概率密度函数为 $f(x)$, 则我们称

$$E(X)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x \tag{3.13}$$

为此随机变量 (或此分布) 的 **均值** 或 **数学期望**. (除下面将要提到的柯西分布以外，有关绝对收敛条件的注意事项与离散型的情况相同.)

下面求曾在第 2 章第 2 节出现过的几个连续分布的均值.

(1) 均匀分布 $\mathrm{U}(a,b)$

$$E(X)=\int_a^b x\frac{1}{b-a}\mathrm{d}x=\frac{a+b}{2}$$

(2) 贝塔分布 $\mathrm{Be}(\alpha,\beta)$

$$\begin{aligned}E(X)&=\int_0^1 x\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x\\&=\frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\\&=\frac{\alpha}{\alpha+\beta}\end{aligned}$$

(3) 指数分布 $\mathrm{Ex}(\alpha)$

$$E(X)=\int_0^{\infty}x\alpha e^{-\alpha x}\mathrm{d}x=1/\alpha$$

(4) 伽玛分布 $\mathrm{G}(\alpha,\nu)$

$$\begin{aligned}E(X)&=\int_0^{\infty}x\frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}\mathrm{d}x \qquad (\text{令}\alpha x=t)\\&=\int_0^{\infty}\frac{1}{\Gamma(\nu)}t^{\nu}e^{-t}\frac{\mathrm{d}t}{\alpha}\\&=\frac{\Gamma(\nu+1)}{\alpha\Gamma(\nu)}=\frac{\nu}{\alpha}\end{aligned}$$

(5) 正态分布 $N(\mu,\sigma^2)$

$$E(X)=\int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]dx \quad (令\frac{x-\mu}{\sigma}=t)$$
$$=\int_{-\infty}^{\infty}(\mu+\sigma t)\frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}t^2]dt$$
$$=\mu\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}t^2]dt+\int_{-\infty}^{\infty}\frac{\sigma t}{\sqrt{2\pi}}\exp[-\frac{1}{2}t^2]dt$$
$$=\mu$$

(6) 柯西分布 $C(\mu,\alpha)$

这是非常典型的不存在均值的分布. 比如在 $\mu=0$, $\alpha=1$ 的情况下

$$E(|X|)=\int_{-\infty}^{\infty}|x|\frac{dx}{\pi(1+x^2)}=\lim_{\substack{u\to\infty\\ l\to-\infty}}\int_{l}^{u}\frac{|x|dx}{\pi(1+x^2)}$$
$$=\lim_{u\to\infty}\int_{0}^{u}\frac{xdx}{\pi(1+x^2)}-\lim_{l\to-\infty}\int_{l}^{0}\frac{xdx}{\pi(1+x^2)}$$
$$=\lim_{u\to\infty}[\frac{1}{2\pi}\log(1+x^2)]_0^u-\lim_{l\to-\infty}[\frac{1}{2\pi}\log(1+x^2)]_l^0$$
$$=\lim_{u\to\infty}\frac{1}{2\pi}\log(1+u^2)+\lim_{l\to-\infty}\frac{1}{2\pi}\log(1+l^2)=\infty$$

同样, 当 μ, α 取其它值时其绝对收敛的条件也不成立. 这是由于当 $|x|$ 趋近无穷大的时候 (与正态分布的概率密度相比较), 概率密度 $f(x)$ 极其缓慢地逼近零. 柯西分布因此也常被称做幅度宽长的分布.

离散分布中所述定理 3-1 及定理 3-3 在这里同样成立.

定理 3-5 设 $g(x)$ 为任意实函数, 则

$$E(g(X))=\int_{-\infty}^{\infty}g(x)f(x)dx \tag{3.14}$$

定理 3-6

$$E(X) = \int_0^{\infty} P(X > x)\mathrm{d}x - \int_{-\infty}^{0} P(X \le x)\mathrm{d}x \tag{3.15}$$

条件期望 设连续型随机变量 X 与 Y 的联合分布的概率密度函数为 $f(x,y)$, Y 的边缘分布的概率密度函数为 $f_Y(y)$. 则对于满足 $f_Y(y) > 0$ 的 Y, 仿离散型的情况，我们同样可以定义 $Y = y$ 条件下 X 分布的概率密度

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} \tag{3.16}$$

$Y = y$ 条件下 X 的数学期望

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f(x|y)\mathrm{d}x \tag{3.17}$$

$Y = y$ 条件下 X 的条件期望与 X 的 (无条件) 数学期望之间具有以下关系.

定理 3-7

$$E(X) = \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y)\mathrm{d}y \tag{3.18}$$

3.1.3 数学期望的性质

如上所述，在计算均值的时候，我们发现离散分布与连续分布有很多共同的性质. 不区分分布的类型，我们将式 (3.1) 及式 (3.13) 表示成下式：

$$E(X) = \int_{-\infty}^{\infty} x\mathrm{d}F(x) \tag{3.19}$$

上式右边为史狄尔切斯 (Stieltjes) 积分 (Stieltjes Integral), 它的具体内容将留给解析学的课本去讨论，这里仅是为了表示方便而已. 当此式出现的时候，如果不便于理解，也可把它看成诸如式 (3.1) 的和的形式，或是诸如式 (3.13) 的黎曼积分 (Riemannian Integral) 的形式.

下面我们讲一些均值的性质. 首先，由均值的定义容易推知其线性性.

[E1] 设 a, b 为任意实常数，则

$$E(aX+b) = aE(X)+b \tag{3.20}$$

设 $X_1, X_2, \cdots, X_n$ 是定义在同一样本空间的随机变量, $g(x_1, x_2, \cdots, x_n)$ 是实函数，则仿定理 3-1 及定理 3-5, 随机变量 $g(X_1, X_2, \cdots, X_n)$ 的数学期望可由下式给出.

[E2] 设 $X_1, X_2, \cdots, X_n$ 的联合分布函数为 $F(x_1, x_2, \cdots, x_n)$, 则

$$\begin{aligned} &E(g(X_1, X_2, \cdots .X_n) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) \mathrm{d}F(x_1, x_2, \cdots, x_n) \end{aligned} \tag{3.21}$$

特别地，当 g 是单变量函数，比如是 x_1 的函数的时候，又有下式成立.

[E3] 设 X_1 的分布函数为 $F_1(x_1)$, 则

$$E(g(X_1)) = \int_{-\infty}^{\infty} g(x_1) \mathrm{d}F_1(x_1) \tag{3.22}$$

从 [E2] 及 [E3], 我们又可以推出 [E4].

[E4] $E(X_1 + X_2 + \cdots + X_n)$

$$= E(X_1) + E(X_2) + \cdots + E(X_n) \tag{3.23}$$

从上式，我们知道各个随机变量之和的数学期望等于各个随机变量的数学期望之和. 这里要注意，它与随机变量 $X_1, X_2, \cdots, X_n$ 是否相互独立无关. 但是在积的情况下，一般来讲，如果没有独立性的话它不能成立.

[E5] 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，则

$$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2)\cdots E(X_n) \tag{3.24}$$

证明 设相互独立随机变量

$X_1, X_2, \cdots, X_n$ 的分布函数为 $F_1, F_2, \cdots, F_n$，则

$$F(x_1, x_2, \cdots, x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n)$$

(第 2 章式 (2.1))，又在 [E2] 中令 $g(x_1, x_2, \cdots, x_n) = x_1 x_2 \cdots x_n$，则

$$\begin{aligned} E(X_1 X_2 \cdots X_n) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1 x_2 \cdots x_n \mathrm{d}F_1(x_1)\mathrm{d}F_2(x_2)\cdots \mathrm{d}F_n(x_n) \\ &= \int_{-\infty}^{\infty} x_1 \mathrm{d}F_1(x_1) \int_{-\infty}^{\infty} x_2 \mathrm{d}F_2(x_2) \cdots \int_{-\infty}^{\infty} x_n \mathrm{d}F_n(x_n) \\ &= E(X_1)E(X_2)\cdots E(X_n) \cdot \end{aligned}$$

[E5] 推广到一般情况，我们可以得到 [E6].

[E6] 设 $g_1(x_1), g_2(x_2), \cdots, g_n(x_n)$ 为任意实函数，并且随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，则

$$\begin{aligned} &E(g_1(X_1)g_2(X_2)\cdots g_n(X_n)) \\ &\quad = E(g_1(X_1)) \quad E(g_2(X_2))\cdots E(g_n(X_n)) \end{aligned} \tag{3.25}$$

3.2 方 差

如果用符号 μ 表示随机变量 X 的数学期望，则 $(X-\mu)^2$ 亦为随机变量，它的均值被称为 X (或 X 分布) 的 **方差**(variance)，并

用 Var(X) 表示.

$$\mathrm{Var}(X) = E\{(X-\mu)^2\} = \int_{-\infty}^{\infty}(x-\mu)^2\mathrm{d}F(x)$$

$$= \begin{cases} \sum_k (v_k-\mu)^2 f(v_k) & (\text{X 服从离散分布}) \\ \int_{-\infty}^{\infty}(x-\mu)^2 f(x)\mathrm{d}x & (\text{X 服从连续分布}) \end{cases} \tag{3.26}$$

此外，我们把 Var(X) 的正的平方根称做 X (或 X 分布) 的标准差 (standard deviation). 当 X 偏离 μ 的概率越来越大时，X 的方差及标准差也越来越大，可见它是一种衡量 X "偏离程度" 的尺度. 比如当 X 服从正态分布 N(μ,σ) 的时候，$\mathrm{Var}(x)=\sigma^2$. 第 2 章图 2-8 给出了 $\sigma=0.5,1,2$ 情况下 (u,σ^2) 分布的概率密度. 表 3-1 还列出了几种具有代表性的分布的方差.

方差具有如下性质：

(1) $\mathrm{Var}(X)\geq 0$ (3.27)

(2) $\mathrm{Var}(X)=E(X^2)-\mu^2$ (3.28)

因为 $(X-\mu)^2=X^2-2\mu X+\mu^2$

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2-2\mu X+\mu^2)=E(X^2)-2E(\mu X)+E(\mu^2)\\ &= E(X^2)-2\mu E(X)+\mu^2\\ &= E(X^2)-2\mu\cdot\mu+\mu^2\\ &= E(X^2)-\mu^2\end{aligned}$$

(3) 设 a,b 为定常数，则

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X) \tag{3.29}$$

因为

$$\begin{aligned}\mathrm{Var}(aX+b) &= E\{(aX+b-(a\mu+b)^2\}=E\{a^2(X-\mu)^2\}\\ &= a^2E\{(x-\mu)^2\}=a^2\mathrm{Var}(X)\end{aligned}$$

(4) 设随机变量 X_1, X_2 的均值为 μ_1, μ_2, 则

$$\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_1)+\mathrm{Var}(X_2)+2E\{(X_1-\mu_1)(X_2-\mu_2)\} \quad (3.30)$$

因为

$$\begin{aligned}
&\mathrm{Var}(X_1 + X_2) = E[\{(X_1 + X_2) - (\mu_1 + \mu_2)\}^2] \\
&= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)] \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2E\{(X_1 - \mu_1)(X_2 - \mu_2)\}
\end{aligned}$$

这里，我们把

$$E\{(X_1 - \mu_1)(X_2 - \mu_2)\}$$

称做随机变量 X_1 与 X_2 的**协方差** (covariance), 并记为 $\mathrm{Cov}(X_1, X_2)$. 另外，又把

$$r(X_1, X_2) = \mathrm{Cov}(X_1, X_2)/\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)} \quad (3.31)$$

称做随机变量 X_1 与 X_2 的**相关系数** (correlation coefficient). 特别地，当 $r(X_1, X_2) = 0$ 的时候，我们称 X_1 与 X_2 不相关. 由于

$$(X_1 - \mu_1)(X_2 - \mu_2) = X_1X_2 - \mu_1X_2 - \mu_2X_1 + \mu_1\mu_1$$

所以在上式的两边取数学期望，则

$$\mathrm{Cov}(X_1, X_2) = E\{(X_1 - \mu_1)(X_2 - \mu_2)\} = E(X_1X_2) - \mu_1\mu_2 \quad (3.32)$$

当 X_1 与 X_2 相互独立的时候，显然有

$$\mathrm{Cov}(X_1, X_2) = r(X_1, X_2) = 0$$

可见 X_1 与 X_2 不相关 (其逆一般不成立).

表 3-1(a) 典型离散型分布一览表

分布名和参数范围	概率分布 $f(k)$ 和变量 k 的范围	均值和方差	特征函数 $\varphi(t)$
(0-1)分布 B$(1;p)$ $0<p<1$ $(q=1-p)$	$p^k q^{1-k}$ $k=0,1$	p pq	$pe^{it}+q$
2项分布 B$(n;p)$ n：自然数 $0<p<1$ $(q=1-p)$	$\binom{n}{k}p^k q^{n-k}$ $k=0,1,\cdots,n$	np npq	$(pe^{it}+q)^n$
几何分布 Ge(p) $0<p<1$ $(q=1-p)$	pq^k $k=0,1,2,\cdots$	q/p q/p^2	$\dfrac{p}{1-qe^{it}}$
负 2 项分布 NB$(r;p)$ r：自然数 $0<p<1$ $(q=1-p)$	$\binom{r+k-1}{k}p^r q^k$ $k=0,1,2,\cdots$	rq/p rq/p^2	$\left(\dfrac{p}{1-qe^{it}}\right)^r$
泊松分布 Po(λ) $\lambda>0$	$e^{-\lambda}\dfrac{\lambda^k}{k!}$ $k=0,1,2,\cdots$	λ λ	$\exp[\lambda(e^{it}-1)]$

注： 表 3-1 中的 $M(a,b,z)$ 为如下合流型超几何级数

$$M(a,b,z)=1+\frac{az}{b}+\frac{(a)_2z^2}{(b)_2 2!}+\frac{(a)_3z^3}{(b)_3 3!}+\cdots,$$
$$(a)_n=a(a+1)(a+2)\cdots(a+n-1)$$

我们把它推广到更一般的情况，对于 n 个随机变量 $X_1,X_2,\cdots,X_n$ 可以得到以下性质.

(5)

$$\mathrm{Var}(X_1,X_2,\cdots,X_n)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}(X_i,X_j) \quad (3.33)$$

特别地，当 $X_1 - X_2 - \cdots - X_n$ 相互独立的时候，有

$$\mathrm{Var}(X_1,X_2,\cdots,X_n)=\mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\cdots+\mathrm{Var}(X_n) \quad (3.34)$$

表 3-1(b) 典型连续型分布一览表

分布名和参数范围	概率分布 $f(k)$ 和变量 k 的范围	均值和方差	特征函数 $\varphi(t)$
均匀分布 $U(a,b)$ $-\infty<a<b<\infty$	$\frac{1}{b-a}$ $a\le x\le b$	$(a+b)/2$ $(b-a)^2/12$	$\frac{e^{ibt}-e^{iat}}{i(b-a)t}$
Beta分布 $Be(\alpha,\beta)$ $\alpha>0,\ \beta>0$	$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ $0\le x\le 1$	$\alpha/(\alpha+\beta)$ $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$M(\alpha,\alpha+\beta,it)$
指数分布 $Ex(\alpha)$ $\alpha>0$	$\alpha e^{-\alpha x}$ $x\ge 0$	$1/\alpha$ $1/\alpha^2$	$\left(1-\frac{it}{\alpha}\right)^{-1}$
Gamma分布 $G(\alpha,\nu)$ $\alpha>0,\ \nu>0$	$\frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}$ $x\ge 0$	ν/α ν/α^2	$\left(1-\frac{it}{\alpha}\right)^{-\nu}$
正态分布 $N(\mu,\sigma^2)$ $-\infty<\mu<\infty,\ \sigma>0$	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ $-\infty<x<\infty$	μ σ^2	$\exp\left[i\mu t-\frac{\sigma^2}{2}t^2\right]$
柯西分布 $C(\mu,\alpha)$ $-\infty<\mu<\infty,\ \alpha>0$	$\frac{1}{\pi}\cdot\frac{\alpha}{(x-\mu)^2+\alpha^2}$ $-\infty<x<\infty$	不存在 不存在	$e^{i\mu t-\alpha\|t\|}$

契比雪夫不等式 (Chebyshev's inequality)　当 $E(X)=\mu$, $\mathrm{Var}(X)=\sigma^2>0$ 的时候，令 $Y=(X-\mu)/\sigma$, 则 $E(Y)=0$, $\mathrm{Var}(Y)=1$. 这种对于给出的随机变量 X 进行形如 $(X-\mu)/\sigma$ 的变换，使其均值变为 0, 方差变为 1 的过程被称做对 X 进行**标准化** (normalization).

设 λ 为任意正数，则 $|Y|\ge\lambda$ 的概率，即 X 与其均值 μ 的偏离程度为标准差 σ 的 λ 倍以上的概率，将随着 λ 的增大而减小．这个概率究竟有多大随 X 的分布而异，但无论 X 服从什么

分布，总有以下不等式成立：

$$P(|X-\mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2} \tag{3.35}$$

这一不等式称做**契比雪夫不等式**.

证明 我们把实数轴分成如下三个区间 $I_1=(-\infty,\mu-\lambda\sigma)$, $I_2=[\mu-\lambda\sigma,\mu+\lambda\sigma]$, $I_3=(\mu+\lambda\sigma,\infty)$, 则 X 的方差为

$$\sigma^2=\int_{I_1}(x-\mu)^2\mathrm{d}F(x)+\int_{I_2}(x-\mu)^2\mathrm{d}F(x)+\int_{I_3}(x-\mu)^2\mathrm{d}F(x)$$

显然在区间 I_1 与 I_3, $|x-\mu| \geq \lambda\sigma$, 在区间 I_2, 上式中的积分为非负，故有

$$\begin{aligned}\sigma^2 &\geq \int_{I_1}(\lambda\sigma)^2\mathrm{d}F(x)+\int_{I_3}(\lambda\sigma)^2\mathrm{d}F(x)\\ &=(\lambda\sigma)^2P(|X-\mu| \geq \lambda\sigma)\end{aligned}$$

成立. 最后在两边除以 $(\lambda\sigma)$, 即可得到式 (3.35).

弱大数定律 (weak law of large numbers) 根据契比雪夫不等式，我们可以推出概率论中极重要的弱大数定律.

定理 3-8 (**弱大数定律**) 设相互独立的随机变量 $X_1,X_2,\cdots,X_n$ 服从同一分布且其均值 μ 相同. 现令 $S_n=X_1+X_2+\cdots+X_n$, 则对于任意 $\varepsilon>0$,

$$\lim_{n\to\infty}P(|\frac{S_n}{n}-\mu| \geq \varepsilon)=0 \tag{3.36}$$

证明 这个定律对于不存在方差的概率分布亦成立，其证明将留给后面的章节去讨论，这里只给出方差 σ^2 存在情况下的证明. 由方差的性质 (3) 及 (5),

$$\mathrm{Var}(\frac{S_n}{n})=\frac{1}{n^2}\mathrm{Var}(S_n)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n} \tag{3.37}$$

另外，由于

$$E(\frac{S_n}{n}) = \mu \tag{3.38}$$

所以根据契比雪夫不等式可以得到

$$P(|\frac{S_n}{n} - \mu| \geq \lambda \frac{\sigma}{\sqrt{n}}) \leq \frac{1}{\lambda^2}$$

令 $\lambda\sigma/\sqrt{n} = \varepsilon$, 则有

$$P(|\frac{S_n}{n} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

于是当 $n \to \infty$ 的时候，得到了式 (3.36).

一般地，设 $Y_1, Y_2, \cdots$ 为一个随机变量序列，如果对于任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P(|Y_n - \mu| \geq \varepsilon) = 0 \tag{3.39}$$

那么我们称序列 $Y_1, Y_2, \cdots$ **依概率收敛**(convergence in probability) 于 μ. 从客观上讲，弱大数定律反映了“在同一分布中独立抽取的样本的算术平均值依概率收敛于这个分布的均值”这样一个性质.

3.3 矩

我们把方差的定义 (3.26) 推广到一般情况，对任意自然数 l, 令

$$\mu_l = E\{(X - \mu)^l\} \tag{3.40}$$

则它是 X 的 l 阶中心矩. 另外，令

$$\mu_l' \equiv E(X^l) \tag{3.41}$$

则它是 X 的 l 阶原点矩.

中心矩与原点矩既可能存在也可能不存在，这将取决于所服从的分布或 l 的取值. 但我们知道，如果存在 l 阶中心矩和 l 阶原点矩，该分布的小于等于 l 阶的中心矩和原点矩一定存在. 反过来，如果不存在 l 阶中心矩和 l 阶原点矩，那么该分布的大于等于 l 阶的中心矩和原点矩也一定不会存在. 中心矩与原点矩之间具有以下关系.

$$\begin{aligned} \mu_2 &= \mu_2' - \mu^2, & \mu_2' &= \mu_2 + \mu^2 \\ \mu_3 &= \mu_3' - 3\mu_2'\mu + 2\mu^3, & \mu_3' &= \mu_3 + 3\mu_2\mu + \mu^3 \\ &\vdots & &\vdots \end{aligned} \tag{3.42}$$

对 X 进行标准化以后得到的关于随机变量 $(X-\mu)/\sigma$ 的三阶及四阶中心矩或原点矩分别为

$$\frac{\mu_3}{\sigma^3}, \qquad \frac{\mu_4}{\sigma^4} \tag{3.43}$$

它被称为 X 分布的**偏度** (skewness) 及**峰度** (kurtosis). 一般地，在正态分布的情况下，它们将变成 0 和 3, 因而，通过 μ_3/σ^3 及 $(\mu_4/\sigma^4)-3$, 我们可以衡量 X 分布与正态分布的偏离程度.

习 题 3

1. 试证 [定理 3-1].

2. 随机变量 X 服从几何分布 $\mathrm{Ge}(p)$, l 为自然数，试求 $Y=\min(X,l)$ 分布的均值.

3. 设 X 为随机变量，c 为任意定常数，试证当 $c=E(X)$ 的时候，$E[(X-c)^2]$ 的值最小.

4. 试证负二项分布 $\mathrm{NB}(r;p)$ 的均值和方差分别为 $rq/p, rq/p^2$ (参照表 3-1). [提示：利用它与几何分布之间的关系.]

5. 设随机变量 X 与 Y 的均值为 0, 方差为 1, 协方差为 ρ. 试证 $Z=X-\rho Y$ 的均值为 0, 方差为 $1-\rho^2$, 且 Z 与 Y 不相关.

6. 试证对于任意随机变量 X 与 Y, 总有

$$E(X^2)E(Y^2)\geq\{E(XY)\}^2$$

成立 (**许瓦尔兹不等式**, Schwarz inequality). 并问在什么情况下等号成立.

[提示: 假定 t 为任意实数, 则 $E\{ctx+r)^2\}\geq 0$, 再考虑其二次判别式的性质.]

7. 根据上一题的结果, 试证任意随机变量 X_1 与 X_2 的相关系数 $r(X_1,X_2)$ 的绝对值小于等于 1. 并问在什么情况下 $r(X_1,X_2)=\pm 1$.

8. 设随机变量 N 只取非负整数值, 且其分布的均值和方差有限. 另外, 设随机变量 $X_1,X_2,\cdots$ 相互独立且服从均值为 μ, 方差为 σ^2 的同一分布, 并且它们与 N 相互独立. 试证"不定个数之和" (定义 $S_0=0$) 的均值和方差如下:

$$E(S_N)=E(N)\mu$$

$$\mathrm{Var}(S_N)=E(N)\sigma^2+\mathrm{Var}(N)\mu^2$$

9. 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且服从均值为 μ, 方差为 σ^2 的同一分布. 现令 $\overline{X}=(X_1+X_2+\cdots+X_n)/n$. 试证:

(1) $\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2$

(2) $\displaystyle E[\sum_{i=1}^{n}(X_i-\overline{X})^2]=(n-1)\sigma^2$

10. 根据契比雪夫不等式, 试求标准化随机变量 Y 的绝对值分别大于等于 1,2,3 的概率的上界. 其次, 求出服从指数分布 Ex(1) 和正态分布 $\mathrm{N}(\mu,\sigma^2)$ 的随机变量的上述概率, 并比较它们的上界.

第 4 章 母函数和特征函数

4.1 概率母函数

设随机变量 X 取非负整数值 $0,1,2,\cdots$，且 $f(k)$ 为其概率分布．对于满足 $-1 \le z \le 1$ 的任一实常数，我们可将随机变量 z^X 的均值 $E(z^X)$ 看作 z 的函数，并把它写成如下形式：

$$G(z) = E(z^X) = \sum_{k=0}^{\infty} f(k) z^k \tag{4.1}$$

当 $|z| \le 1$ 的时候，

$$|G(z)| \le \sum_{k=0}^{\infty} f(k)|z|^k \le \sum_{k=0}^{\infty} f(k) = 1$$

级数 (4.1) 收敛，且此时可逐项进行任意次微分．比如，

$$G'(z) = \sum_{k=1}^{\infty} f(k) k z^{k-1} = E(X z^{X-1}) \tag{4.2}$$

$$G''(z) = \sum_{k=2}^{\infty} f(k) k(k-1) z^{k-2} = E\{X(X-1) z^{X-2}\} \tag{4.3}$$

一般地，

$$\begin{aligned} G^{(l)}(z) &= \sum_{k=l}^{\infty} f(k) k(k-1)\cdots(k-l+1) z^{k-l} \\ &= E\{X(X-1)\cdots(X-l+1) z^{X-l}\} \end{aligned} \tag{4.4}$$

这里，令 $z=0$, 则

$$G^{(l)}(0) = f(l)l! \Rightarrow f(l) = G^{(l)}(0)/l! \tag{4.5}$$

式 (4.1) 表示由概率分布 $f(k)$ 可以确定 $G(z)$. 反过来，式 (4.5) 表示由 $G(z)$ 可以确定概率分布 $f(k)$. 我们称 $G(z)$ 为 **概率母函数** (probability generating function). 概率分布与概率母函数之间具有一一对应关系.

在式 (4.2) 及式 (4.3) 中，令 $z=1$, 则

$$G'(1) = E(X) \tag{4.6}$$

$$G''(1) = E\{X(X-1)\} = E(X^2) - E(X) \tag{4.7}$$

因此,

$$E(X) = G'(1) \tag{4.8}$$

$$\mathrm{Var}(X) = E(X^2) - E\{(X)\}^2 = G''(1) + G'(1) - \{G'(1)\}^2 \tag{4.9}$$

可见，某一分布的均值及方差均可由概率母函数算出.

例 4-1 当随机变量 X 服从二项分布的时候,

$$f(k) = \begin{cases} \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-q)^{n-k} & (0 \le k \le n) \\ 0 & (k > n) \end{cases}$$

因而,

$$G(z) = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}z^k = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} (pz)^k(1-p)^{n-k}$$
$$= (pz+q)^n \qquad (q = 1-p) \tag{4.10}$$

由此，我们得到

$$G'(1) = np(pz+q)^{n-1}|_{z=1} = np = E(X)$$
$$G''(1) = n(n-1)p^2(pz+q)^{n-2}|_{z=1} = n(n-1)p^2$$
$$\text{Var}(X) = n(n-1)p^2 + np - (np)^2 = npq$$

例 4-2 当随机变量 X 服从泊松分布 $\text{Po}(\lambda)$ 的时候，

$$f(k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad (k \geq 0)$$

因而，

$$G(z) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}z^k = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda z)^k}{k!} = e^{-\lambda}\cdot e^{\lambda z} = e^{\lambda(z-1)} \quad (4.11)$$

对它求导，则有

$$G'(z) = \lambda e^{\lambda(z-1)}, \qquad G''(z) = \lambda^2 e^{\lambda(z-1)}$$

由此，我们得到

$$E(X) = \lambda, \qquad \text{Var}(x) = \lambda$$

通过概率母函数可以方便地求出多个相互独立的非负整数值随机变量的和的分布．这里，要用到下面的性质．

定理 4-1 设非负整数值随机变量 $X_1, X_2, \cdots, X_r$ 相互独立，且

$$S = X_1 + X_2 + \cdots + X_r$$

假定 $G_i(z)$ 为 X_i 的概率母函数，$G_s(z)$ 为 S 的概率母函数，则

$$G_s(z) = G_1(z)G_2(z)\cdots G_r(z) \quad (4.12)$$

证明 当 $r=2$ 的时候，如果能够证明上式成立，则对于一般情况，我们可以通过数学归纳法推出. 下面仅就 $r=2$ 的情况进行证明.

根据第 2 章第 5 节的内容，我们知道

$$\begin{aligned}P(X_1+X_2=k) &= \sum_{l=0}^{k} P(X_1=l \text{ 且 } X_2=k-l) \\ &= \sum_{l=0}^{k} P(X_1=l)P(X_2=k-l)\end{aligned}$$

因此，

$$\begin{aligned}G_s(z) &= \sum_{k=0}^{\infty} P(X_1+X_2=k)z^k \\ &= \sum_{k=0}^{\infty}\{\sum_{l=0}^{k} P(X_1=l)P(X_2=k-l)z^l \cdot z^{k-l}\}\end{aligned}$$

交换一下取和的顺序，并令 $k-l$ 为 k', 则

$$\begin{aligned}G_s(z) &= \sum_{l=0}^{\infty} P(X_1=l)z^l \sum_{k'=0}^{\infty} P(X_2=k')z^{k'} \\ &= G_1(z)G_2(z)\end{aligned}$$

例 4-3 设相互独立的随机变量 $X_1, X_2, \cdots, X_r$ 分别服从二项分布 $\mathrm{B}(n_i;p)$, 那么 $S=X_1+X_2+\cdots+X_r$ 也服从二项分布 $\mathrm{B}(n_1+n_2+\cdots+n_r;p)$. 这是因为，由例 4-1,

$$G_i(z) = (pz+1-p)^{n_i}$$

所以

$$G_s(z) = (pz+1-p)^{n_1+n_2+\cdots+n_r}$$

而它即为二项分布 $B(n_1+n_2+\cdots+n_r;p)$ 的概率母函数.

问题 设相互独立的随机变量 $X_1,X_2,\cdots,X_r$ 分别服从泊松分布 $Po(\lambda_i)$, 试证 $S=X_1+X_2+\cdots+X_r$ 也服从泊松分布 $Po(\lambda_1+\lambda_2+\cdots+\lambda_r)$.

这里所反映的二项分布及泊松分布的性质, 即"服从某一分布 (参数可以不同) 的多个相互独立随机变量的和的分布, 同样服从这个分布"的性质, 被称做分布的**再生性** (reproductive of a distribution). 除了这两个分布以外, 我们还知道另外一些具有再生性的分布.

4.2 矩母函数

设 X 为任意随机变量 (连续型或者离散型), θ 为实定常数. 则 $e^{\theta X}$ 亦为随机变量, 其均值为

$$M(\theta)\equiv E(e^{\theta X})=\int_{-\infty}^{\infty}e^{\theta x}\mathrm{d}F(x) \tag{4.13}$$

对于包含 $\theta=0$ 在内的适当范围的 θ, 如果均值存在 (即取有限值), 我们把 $M(\theta)$ 看做 θ 的函数, 并称它为**矩母函数** (moment generating function). $M(\theta)$ 在 $\theta=0$ 处可进行任意次微分, 并且

$$M^{(l)}(0)=E(X^l)=\mu'_l \tag{4.14}$$

可见通过对 $M(\theta)$ 求导, 我们可以求出任意阶的原点矩. 矩母函数的名称, 就来自这个性质.

当 X 的矩母函数为 $M(\theta)$ 的时候, $Y=aX+b$ (a, b 为定常数) 的矩母函数为

$$E(e^{\theta(aX+b)})=E(e^{a\theta X}\cdot e^{b\theta})=e^{b\theta}M(a\theta) \tag{4.15}$$

另外，当随机变量 X 只取非负整数值的时候，矩母函数 $M(\theta)$ 与概率母函数 $G(z)$ 之间具有以下关系.

$$M(\theta)=G(e^{\theta}) \tag{4.16}$$

例 4-4 设 X 服从标准正态分布 N(0, 1), 则

$$\begin{aligned} M(\theta) &= \int_{-\infty}^{\infty} e^{\theta x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\mathrm{d}x \\ &= e^{\theta^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(x-\theta)^2/2}\mathrm{d}x = e^{\theta^2/2} \end{aligned} \tag{4.17}$$

在 $\theta=0$ 附近进行泰勒展开，它变为

$$M(\theta)=\sum_{k=0}^{\infty}\frac{(\theta^2/2)^k}{k!}=\sum_{k=0}^{\infty}\frac{\theta^{2k}}{2^k k!}$$

由此可知，这个分布的奇数阶原点矩为 0, $2k$ 阶原点矩为 $(2k)!/(2^k k!)$.

此外，我们还知道 $Y=\sigma X+\mu$ 服从正态分布 $\mathrm{N}(\mu,\sigma)$, 其矩母函数由式 (4.15) 为

$$e^{\mu\theta}\cdot e^{(\sigma\theta)^2/2}=\exp[\mu\theta+\frac{\sigma^2}{2}\theta^2] \tag{4.18}$$

例 4-5 在概率母函数 $G(z)=(pz+1-p)^n$ 中用 e^{θ} 代替 z, 然后再通过式 (4.16), 我们便可以得到二项分布 $\mathrm{B}(n;p)$ 的矩母函数.

$$M(\theta)=(pe^{\theta}+1-p)^n \tag{4.19}$$

将此式关于 θ 求导,

$$M'(\theta)=npe^{\theta}(pe^{\theta}+1-p)^{n-1}$$

$$M''(\theta)=npe^{\theta}(pe^{\theta}+1-p)^{n-1}+n(n-1)(pe^{\theta})^2(pe^{\theta}+1-p)^{n-2}$$

由此可以推出这个分布的均值和方差

$$E(X) = M'(0) = np$$
$$\mathrm{Var}(X) = M''(0) - \{E(X)\}^2 = np(1-p)$$

关于相互独立随机变量的和的矩母函数，仿定理 4-1, 有以下定理.

定理 4-2 设相互独立的随机变量 $X_1, X_2, \cdots, X_r$ 的矩母函数分别为 $M_1(\theta), M_2(\theta), \cdots, M_r(\theta)$, $S = X_1 + X_2 + \cdots + X_r$ 的矩母函数为 $M_s(\theta)$, 则

$$M_s(\theta) = M_1(\theta)M_2(\theta)\cdots M_r(\theta) \tag{4.20}$$

4.3 特征函数

并非所有分布都存在矩母函数. 比如，不存在均值的柯西分布就不存在矩母函数. 在式 (4.13) 中，我们用 $\mathrm{i}t$ 来替换 z (t 为实数，i 为虚数单位 $\sqrt{-1}$), 则

$$\varphi(t) \equiv E(\mathrm{e}^{\mathrm{i}tX}) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x) \tag{4.21}$$

我们称此式为随机变量 X (或 X 分布) 的**特征函数**. 由于

$$|\varphi(t)| \leq E(|e^{\mathrm{i}tX}|) = E(1) = 1$$

所以任何分布都存在特征函数.

当矩母函数 $M(\theta)$ 存在的时候，若用 $\mathrm{i}t$ 代替 θ, 可以得到 $\varphi(t)$.

$$\varphi(t) = M(\mathrm{i}t) \tag{4.22}$$

比如，

$$\text{对于二项分布 } \mathrm{B}(n;p): \quad \varphi(t) = (pe^{\mathrm{i}t} + 1 - p)^n \tag{4.23}$$

$$\text{对于正态分布 } \mathrm{N}(\mu,\sigma^2): \quad \varphi(t) = \exp(\mathrm{i}\mu t - \frac{\sigma^2}{2}t^2) \tag{4.24}$$

另一方面，通过对特征函数求导可以推出矩. 即由

$$\varphi^{(l)}(t) = E\{(\mathrm{i}X)^l e^{\mathrm{i}tX}\}, \quad \varphi^{(l)}(0) = E\{(\mathrm{i}X)^l\}$$

可以得到

$$E(X^l) = \varphi^l(0)/\mathrm{i}^l \tag{4.25}$$

特别地，我们还可以得到

$$E(X) = \varphi'(0)/\mathrm{i} \tag{4.26}$$

$$\mathrm{Var}(X) = -\varphi''(0) + \{\varphi'(0)\}^2 \tag{4.27}$$

如果存在直至 s 阶的原点矩 μ_l' $(0 \le l \le s)$, 我们可以在 $t = 0$ 的附近对 $\varphi(t)$ 进行泰勒展开，即

$$\varphi(t) = 1 + \sum_{l=1}^{s} \frac{(\mathrm{i}t)^l}{l!} u_l' + o(t^s) \tag{4.28}$$

这里，$o(t^s)$ 为 t^s 的高阶无穷小，即

$$\lim_{t\to 0} \frac{o(t^s)}{t^s} = 0$$

对于特征函数，仿定理 4-1 及定理 4-2, 我们得到如下定理.

定理 4-3 对于相互独立的随机变量 $X_1, X_2, \cdots, X_r$, 设 X_j 的特征函数为 $\varphi_j(t)$, 则 $S = X_1 + X_2 + \cdots + X_r$ 的特征函数 $\varphi_s(t)$ 为

$$\varphi_s(t) = \varphi_1(t)\varphi_2(t)\cdots\varphi_r(t) \tag{4.29}$$

证明

$$\begin{aligned}\varphi_s(t) &= E(e^{\mathrm{i}t(X_1+X_2+\cdots+X_r)}) = E(e^{\mathrm{i}tX_1}\cdot e^{\mathrm{i}tX_2}\cdot\cdots\cdot e^{\mathrm{i}tX_r}) \\ &= E(e^{\mathrm{i}tX_1})E(e^{\mathrm{i}tX_2})\cdots E(e^{\mathrm{i}tX_r}) \\ &= \varphi_1(t)\varphi_2(t)\cdots\varphi_r(t)\end{aligned}$$

后面我们将要讲到，这个定理对于考察随机变量和的性质起重要作用.

特征函数是考察概率分布性质的重要工具. 正如下面定理所述，其理由在于概率分布与特征函数具有一一对应关系，且如果两个概率分布相近，它们的特征函数也彼此接近.

定理 4-4 设随机变量 X_1, X_2 的分布函数分别为 $F_1(x), F_2(x)$，特征函数分别为 $\varphi_1(t), \varphi_2(t)$，那么 $F_1(x) \equiv F_2(x)$ 的充要条件是 $\varphi_1(t) = \varphi_2(t)$.

特别地，对于整数型随机变量，有如下的 **逆转公式** (inversion formula)：

$$f(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathrm{i}kt}\varphi(t)\mathrm{d}t \tag{4.30}$$

对于连续型随机变量，也有如下的逆转公式：

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}xt}\varphi(t)\mathrm{d}t \tag{4.31}$$

定理 4-5 连续定理(continuity theorem)

设 $F_1(x), F_2(x), \cdots$ 为分布函数序列， $\varphi_1(t), \varphi_2(t), \cdots$ 为对应的特

征函数序列．对于任意实数 t, 如果数列 $\varphi_1(t),\varphi_2(t),\cdots$ 收敛，即

$$\lim_{n\to\infty}\varphi_n(t)=\varphi(t) \tag{4.32}$$

且 $\varphi(t)$ 在 $t=0$ 连续，则 $\varphi(t)$ 为特征函数．此外，设 $F(x)$ 对应于 $\varphi(t)$ 的分布函数，则对于 $F(x)$ 的所有连续点

$$\lim_{n\to\infty}F_n(x)=F(x) \tag{4.33}$$

例 4-6(正态分布，伽玛分布的再生性) 如果相互独立的随机变量 $X_k(k=1,2,\cdots,n)$ 分别服从正态分布 $\mathrm{N}(\mu_k,\sigma_k^2)$, 由式 (4.24) 及定理 4-3, 它们的和 S 分布的特征函数 $\varphi_s(t)$ 为

$$\varphi_s(t)=\prod_{k=1}^{n}\exp[\mathrm{i}\mu_k t-\frac{\sigma_k^2}{2}t^2]=\exp[\mathrm{i}\mu t-\frac{\sigma^2}{2}t^2]$$

$$\sigma^2=\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2$$

$$\mu=\mu_1+\mu_2+\cdots+\mu_n$$

即 S 服从正态分布 $\mathrm{N}(\mu,\sigma^2)$.

又，伽玛分布 $G(\alpha,\nu)$ 的特征函数[1]

$$\begin{aligned}\varphi(t)&=\int_0^\infty e^{\mathrm{i}tx}\frac{1}{\Gamma(\nu)}\alpha^\nu x^{\nu-1}e^{-\alpha x}\mathrm{d}x\\&=(1-\frac{\mathrm{i}t}{\alpha})^{-\nu}\end{aligned} \tag{4.34}$$

[1] 为了求出这个特征函数，严格地讲需要有关复数积分的知识．但在这里，我们不去深入地考察它，而仅通过形式上的变量代换来推出式 (4.34)．令 $(\alpha-\mathrm{i}t)x=y$, 然后进行积分变换，则有

$$\varphi(t)=\int_0^\infty\frac{1}{\Gamma(\nu)}(\frac{\alpha}{\alpha-\mathrm{i}t})^\nu e^{-y}y^{\nu-1}dy=(\frac{\alpha}{\alpha-\mathrm{i}t})^\nu=(1-\frac{\mathrm{i}t}{\alpha})^{-\nu}$$

如果相互独立的随机变量 $X_k(k=1,2,\cdots,n)$ 分别服从伽玛分布 $\mathrm{G}(\alpha,\nu_k)$，它们的和 S 分布的特征函数 $\varphi_s(t)$ 为

$$\varphi_s(t)=\prod_{k=1}^{n}(1-\frac{\mathrm{i}t}{\alpha})^{-\nu_k}=(1-\frac{\mathrm{i}t}{\alpha})^{-\nu},\quad \nu=\nu_1+\nu_2+\cdots+\nu_n$$

即 S 服从伽玛分布 $\mathrm{G}(\alpha,\nu)$.

利用特征函数，非常容易判定两个随机变量是否相互独立.

首先，我们定义随机变量 $X_1,X_2,\cdots,X_r$ 的联合分布特征函数为

$$\varphi(t_1,t_2,\cdots,t_r)=E\{\exp[\mathrm{i}(t_1X_1+t_2X_2+\cdots+t_rX_r)]\} \tag{4.35}$$

又令每个随机变量 X_j 的特征函数为 $\varphi_j(t_j)$，那么很显然，

$$\varphi_1(t_j)=\varphi(t_j,0,\cdots,0) \tag{4.36}$$

于是对应于第 2 章的式 (2.51)，有以下定理.

定理 4-6 随机变量 $X_1,X_2,\cdots,X_r$ 相互独立的充要条件是

$$\varphi(t_1,t_2,\cdots,t_r)\equiv\varphi_1(t_1)\varphi_2(t_2)\cdots\varphi_r(t_r) \tag{4.37}$$

例 4-7 设相互独立的随机变量 Y_1,Y_2 分别服从同一正态分布 $N(\mu,\sigma^2)$，试证 $X_1=Y_1+Y_2$，$X_2=Y_1+Y_2$ 相互独立.

首先，我们知道 X_1 与 $X2$ 联合分布的特征函数为

$$\begin{aligned}
\varphi(t_1,t_2) =& E\{\exp[\mathrm{i}t_1(Y_1+Y_2)+\mathrm{i}t_2(Y_1-Y_2)]\}\\
=& E\{\exp[\mathrm{i}(t_1+t_2)Y_1+\mathrm{i}(t_1-t_2)Y_2]\}\\
=& E\{\exp[\mathrm{i}(t_1+t_2)Y_1]\}\cdot E\{\exp[\mathrm{i}(t_1-t_2)Y_2]\}\\
=& \exp[\mathrm{i}\mu(t_1+t_2)-\frac{\sigma^2}{2}(t_1+t_2)^2]\\
& \times\exp[\mathrm{i}\mu(t_1-t_2)-\frac{\sigma^2}{2}(t_1-t_2)^2]\\
=& \exp[2\mathrm{i}\mu t_1-\sigma^2(t_1^2+t_2^2)]
\end{aligned}$$

另一方面，X_1 服从正态分布 $\mathrm{N}(2\mu,2\sigma^2)$, X_2 服从正态分布 $N(0,2\sigma^2)$, 故它们的特征函数分别为

$$\varphi_1(t_1)=\exp[\mathrm{i}(2\mu)t_1-\frac{2\sigma^2}{2}t_1^2]$$
$$\varphi_2(t_2)=\exp[\mathrm{i}(0)t_2-\frac{2\sigma^2}{2}t_2]$$

可见，$\varphi(t_1,t_2)=\varphi_1(t_1)\varphi_2(t_2)$, 再由定理 4-6 可知 X_1 与 X_2 相互独立.

根据连续定理 (定理 4-5), 不必假定方差的存在即可推出第 3 章叙述过的弱大数定律 (定理 3-8). 设相互独立的随机变量 $X_1,X_2,\cdots$ 服从同一分布，且此分布的均值 (有限值) 为 μ, 特征函数为 $\varphi(t)$. 如果令 $S_n/n=(X_1+X_2+\cdots+X_n)/n$ 的特征函数为 $\varphi_n(t)$, 则

$$\varphi_n(t)=[\varphi(t/n)]^n \tag{4.38}$$

仿式 (4.28), 对 $\varphi(t)$ 进行泰勒展开，则

$$\varphi(t)=1+\mathrm{i}\mu t+o(t)$$

另一方面，在 $t=0$ 附近对 $e^{\mathrm{i}\mu t}$ 进行泰勒展开，则

$$e^{\mathrm{i}\mu t}=1+\mathrm{i}\mu t+o(t)$$

故

$$\varphi(t)=e^{\mathrm{i}\mu t}+o(t)$$

最后，我们推出

$$\varphi_n(t)=[e^{\mathrm{i}\mu t/n}+o(t/n)]^n\to e^{\mathrm{i}\mu t}\qquad(n\to\infty)$$

因为 $e^{\mathrm{i}\mu t}$ 是取 μ 值的随机变量的特征函数，所以由连续定理，我们知道对于任意 $\varepsilon>0$, 都有

$$\lim_{n\to\infty}P(|\frac{s_n}{n}-\mu|>\varepsilon)=0$$

4.4 中心极限定理

如前所述，如果相互独立的随机变量 $X_1,X_2,\cdots,X_n$ 服从均值为 μ、方差为 σ^2 的同一分布，它们的和 $S_n=X_1+X_2+\cdots+X_n$ 在应用上具有非常重要的意义. 一方面，我们知道大数定律反映 $\frac{S_n-n\mu}{n}$ 在 $n\to\infty$ 的时候收敛于 0, 即

$$\frac{S_n-n\mu}{n}=\frac{S_n}{n}-\mu\to 0\quad\text{(依概率收敛)}\tag{4.39}$$

另一方面，我们还知道 $S_n-n\mu$ 的均值为 0, 方差为 $n\sigma^2$, 因而在 $n\to\infty$ 的时候，方差趋于无穷大，即不收敛. 现在，让我们考察一下介于 $(S_n-n\mu)/n$ 与 $(S_n-n\mu)/1$ 之间的“中间量”

$$S_n^*=(S_n-n\mu)/(\sqrt{n}\sigma)\tag{4.40}$$

很显然，S_n^* 是对 S_n 标准化的结果，因而对于任意 n, 其均值都为 0, 方差都为 1. 除此之外，它还有一个明显的特征，即无论 $X_1,X_2,\cdots$

的分布如何，S_n^* 的分布近似地服从标准正态分布 N(0,1). 也就是说对于任意实数 a^*, 都有

$$\lim_{n\to\infty} P(S_n^* \le a^*) = \Phi(a^*) = \int_{-\infty}^{a^*} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x \tag{4.41}$$

这就是在概率论和统计学中起主导作用的 **中心极限定理** (central limit theorem).

在图 4-1 中，我们将 $X_1, X_2, \cdots, X_{10}$ 服从 0-1 分布时 S_{10}^* (即二项分布 B(10;0.3) 的标准化形式) 的概率分布与标准正态分布的概率密度做了比较. 很显然，中心极限定理反映的是 n 趋近于无穷大时的情况，但 $n = 10$ 的时候，我们发现 S_n^* 已相当接近标准正态分布.

图 4-2 给出了均匀分布 U(0,1) 的 n 个相互独立随机变量之和 S_n 的标准化形式 S_n^* 的概率密度函数 $f_n(x)$. $n = 6$ 的时候，其曲线已相当接近标准正态分布的 (参照图 4-1).

由此可见，n 的取值即使不是很大，也总有近似公式

$$P(S_n \le a) = P(S_n^* \le a^*) \approx \Phi(a^*), \quad a^* = \frac{a - n\mu}{\sqrt{n}\sigma} \tag{4.42}$$

成立 [1]. 利用这个公式，我们可以通过简单地正态分布表比较精确地求出那些难于正确计算的概率.

---

[1] 当 $X_1, X_2, \cdots$ 服从连续分布的时候，可用 $S_n < a$ 来代替这个近似公式中的 $S_n \le a$. 当 S_n 服从只取整数值的离散分布的时候，假定 a 为整数，则有

$$p(S_n \le a) = P(S_n \le a + 0.5)$$

因而用

$$P(S_n \le a) \approx \Phi(\frac{a + 0.5 - n\mu}{\sqrt{n}\sigma}) \tag{4.43}$$

来代替式 (4.42) 更为有效. 这个补正被称为 **不连续补正** (discontinuity correction) (有时也被称为连续补正，continuity correction). n 的取值较小的时候，它非常有效.

例 4-8 试计算服从二项分布 B(50; 0.2) 的随机变量 S_{50} 的值小于等于 15 的概率 $P(S_{50} \leq 15)$.

正确地讲，这个概率需要通过计算下式

$$\sum_{k=0}^{15} \binom{50}{k} (0.2)^k (0.8)^{50-k}$$

得出．这会非常麻烦．但由于

$$E(S_{50}) = 50 \times 0.2 = 10, \qquad \mathrm{Var}(S_{50}) = 50 \times 0.2 \times 0.8 = 8$$

我们可以利用近似公式得到

$$\begin{aligned} P(S_{50} \leq 15) &\fallingdotseq P(S_{50}^* \leq (15-10)/\sqrt{8}) \\ &\approx \Phi(1.768) \\ &\approx 0.9614 \end{aligned}$$

(如果采用不连续补正， $\Phi(15.5-10)/\sqrt{8} \approx 0.9741$. 其精确值近似地等于 0.9692).

例 4-9 设随机变量 $X_1, X_2, \cdots, X_{20}$ 服从指数分布 Ex(0.5), 试求 $P(30 \leq S_{20} \leq 50)$.

$$\begin{aligned} E(S_{20}) = 20/0.5 = &\ 40, \qquad \mathrm{Var}(S_{20}) = 20/(0.5)^2 = 80 \\ P(30 \leq S_{20} \leq 50) \fallingdotseq &\ P(\frac{30-40}{\sqrt{80}} \leq S^* \leq \frac{50-40}{\sqrt{80}}) \\ \approx &\ \Phi(1.118) - \Phi(-1.118) \\ \approx &\ 0.7364 \end{aligned}$$

中心极限定理的证明[1] 可利用特征函数证明.

[1] 关于中心极限定理，有许多参考书．比如，清水良一的《中心极限定理》(新应用数学系列 14, 日本教育出版社， 1976).

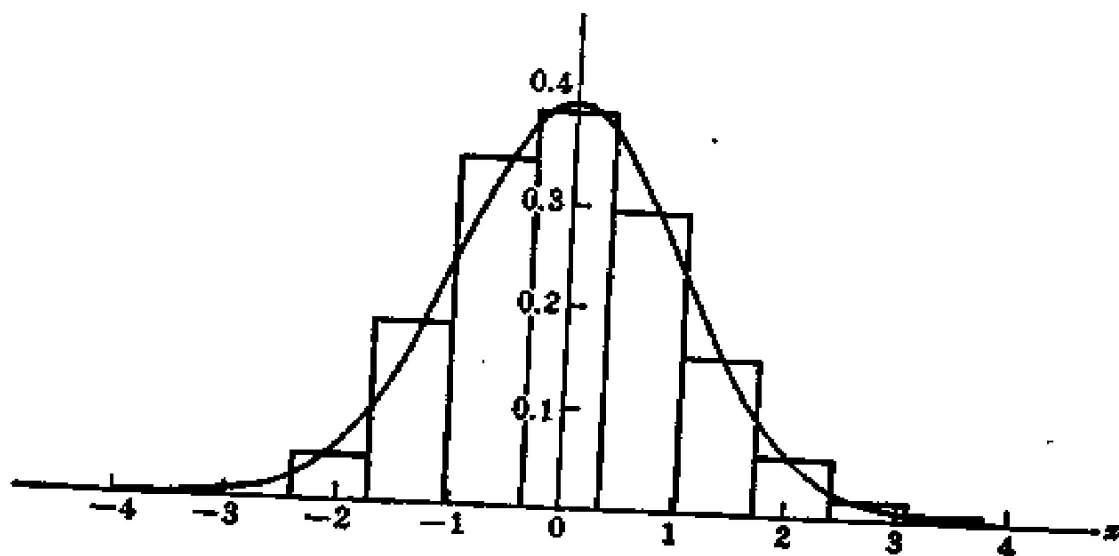


图 4-1 将二项分布 B(10;0.3) 的随机变量标准化以后得到的 (S_{10}^*) 的概率分布与标准正态分布的概率密度函数比较图，(S_{10}^*) 的取值等于"柱"的(横向中心的 x 坐标，而其概率等于柱的面积.

设 $S_n^* = \sum_{j=1}^{n}(X_j - \mu)/\sqrt{n}\sigma$ 的特征函数为 $\varphi_n(t)$, $(X_j-\mu)/(\sqrt{n}\sigma)$ 的特征函数为 $\varphi(t)$, 则由定理 4-6,

$$\varphi_n(t) = [\varphi(t)]^n$$

另一方面，由式 (4.28),

$$\varphi(t) = 1 - \frac{t^2}{2n} + o(\frac{t^2}{n})$$

因此，我们可以得到

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty}[1 - \frac{t^2}{2n} + o(\frac{t^2}{n})]^n = e^{-t^2/2}$$

由于 $e^{-t^2/2}$ 为标准正态分布的特征函数，通过连续定理可以推出 S_n^* 的分布收敛于标准正态分布.

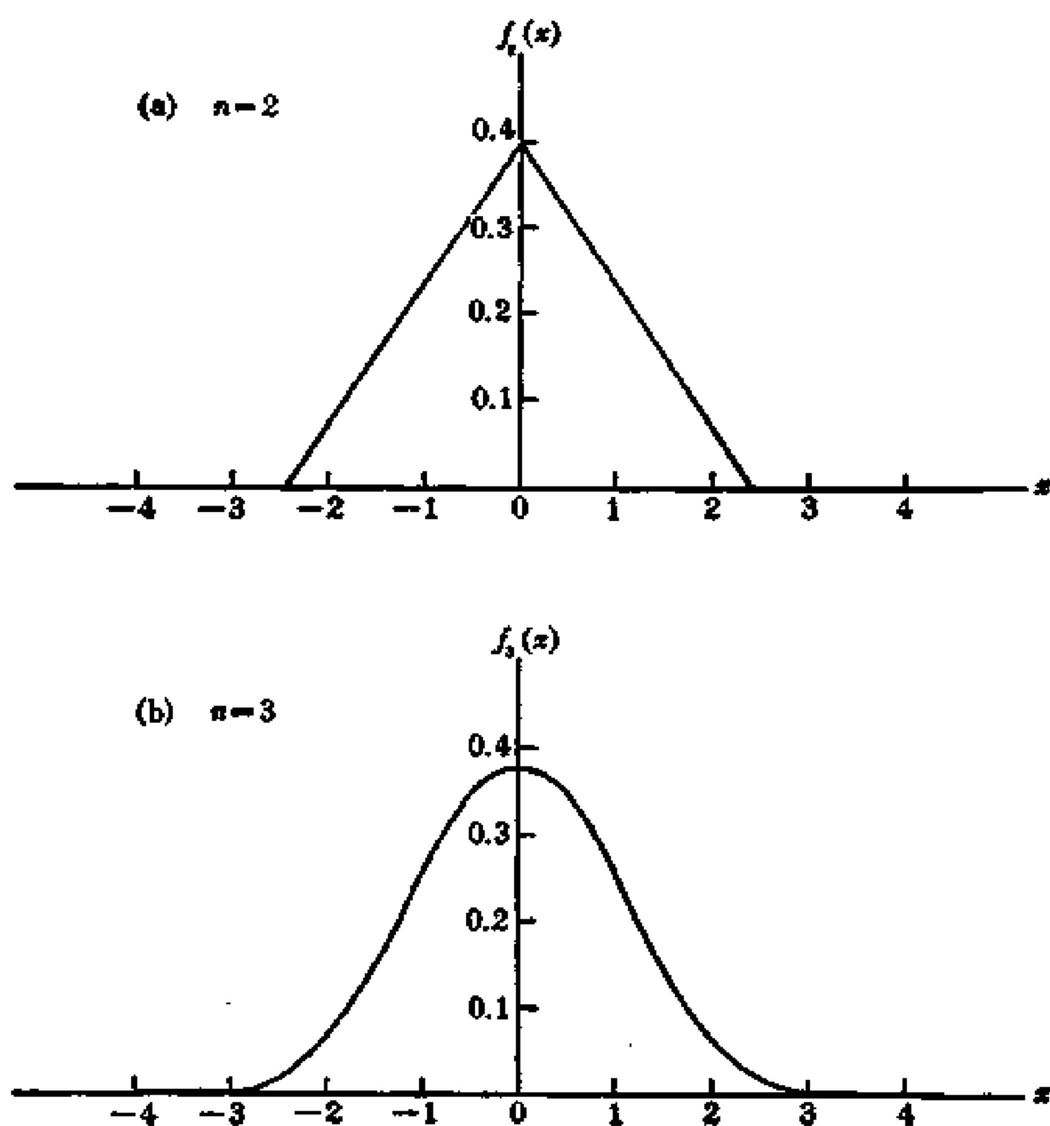
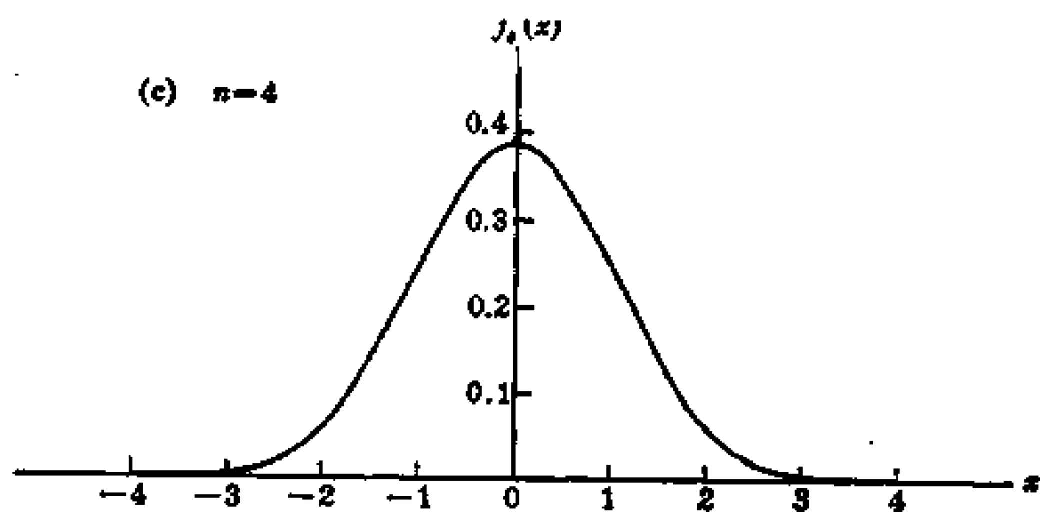
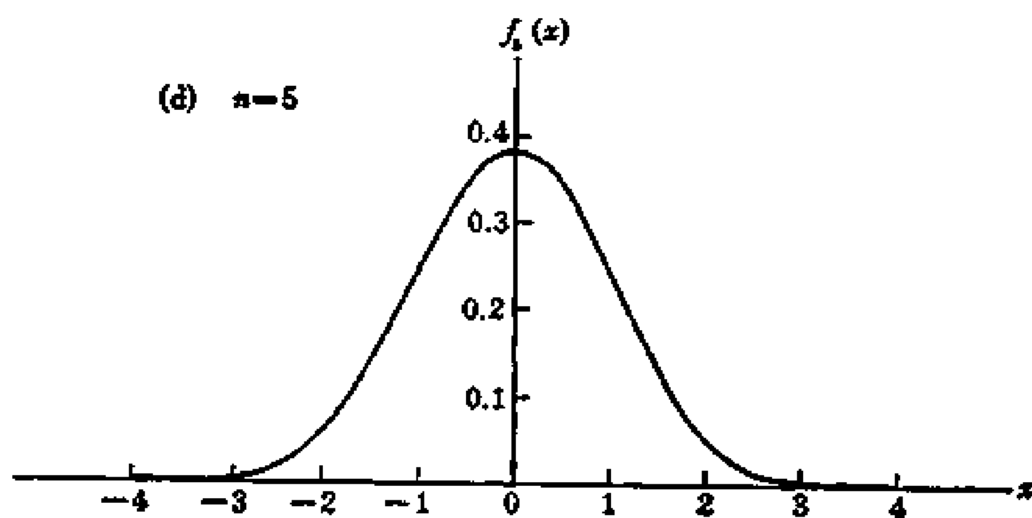


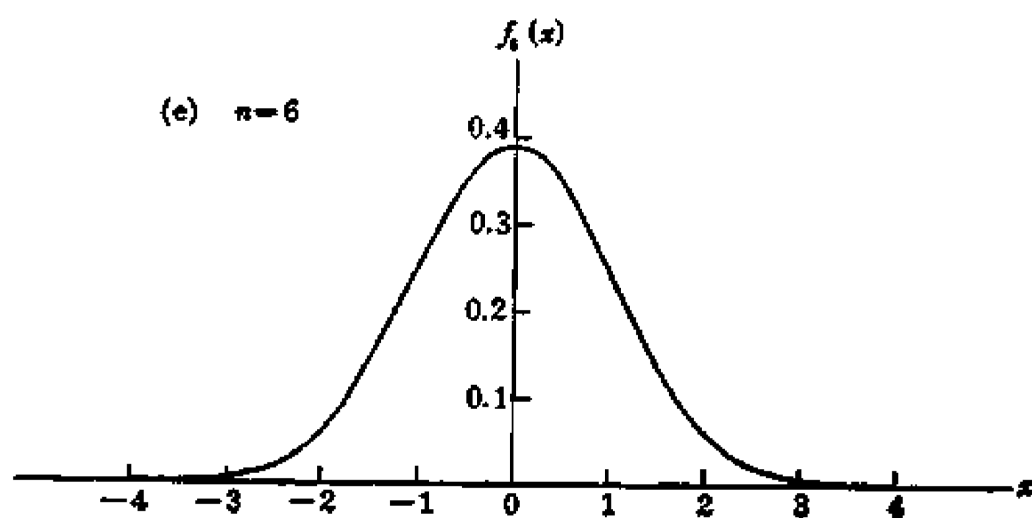
图 4-2　将 n 个相互独立地服从均匀分布 U(0;1) 的随机变量之和标准化以后得到的分布的概率密度函数 $f_n(x)$



$f_4(x)$
(c) $n=4$
0.4
0.3
0.2
0.1
−4
−3
−2
−1
0
1
2
3
4
x



$f_5(x)$
(d) $n=5$
0.4
0.3
0.2
0.1
−4
−3
−2
−1
0
1
2
3
4
x



$f_6(x)$
(e) $n=6$
0.4
0.3
0.2
0.1
−4
−3
−2
−1
0
1
2
3
4
x

图 4-2 （续）

习 题 4

1.(1) 试求几何分布 Ge(p) 的概率母函数.

(2) 利用 (1) 的结果，试证此分布的均值和方差分别为 q/p, q/p^2 ($q = 1 - p$).

(3) 我们已经知道，相互独立地服从几何分布 Ge(p) 的 r 个随机变量之和的分布服从负二项分布 NB($r;p$). 利用这个条件，试求 NB($r;p$) 的概率母函数.

(4) 试证负二项分布具有再生性.

(5) 令 $rq = \lambda$(定常数). 试证 q 趋近于零的时候，NB($r;p$) 的概率母函数趋近于泊松分布的概率母函数.

2. 利用中心极限定理，试求下面概率的近似值.

(1) 当 X 服从泊松分布 Po(10) 的时候，$P(X \leq 12)$.

(2) 当 X 服从负二项分布 NB(20;0.2) 的时候，$P(X \geq 100)$.

(3) 当 X 服从伽玛分布 G(1,40) 的时候，$P(X \geq 50)$.

并说出可以根据中心极限定理算出它们概率近似值的理由.

3. 一般来讲，对于实数列 $\{a_k; k = 0, 1, 2, \cdots\}$, 如果实变量 z 的级数

$$A(z) = a_0 + a_1 z + a_2 z^2 + \cdots$$

在适当范围 $-z_0 < z < z_0$ 内收敛，我们称 $A(z)$ 为此数列的**母函数** (generating function). 概率母函数是满足 $a_k \geq 0$, $\sum_{k=0}^{\infty} a_k = 1$ 的数列的母函数.

(1) 设服从概率分布 $\{f(k); k = 0, 1, 2, \cdots\}$ 的随机变量 X 的概率母函数为 $G(z)$. 当

$$q_k = P(X > k)$$

的时候，试证数列 $\{q_k; k=0,1,2,\cdots\}$ 的母函数 $Q(z)$ 为

$$Q(z)=\frac{1-G(z)}{1-z} \tag{4.44}$$

(2) 试证

$$E(X)=Q(1), \qquad \mathrm{Var}(X)=2Q'(1)+Q(1)-Q^2(1) \tag{4.45}$$

这里，$Q(1)$ 和 $Q'(1)$ 分别为 z 从小于 1 的方向趋近于 1 时，$Q(z)$ 与 $Q'(z)$ 的极限值.

4. 掷 k 次骰子，假定"1"出现偶数次的概率为 a_k.

(1) 试证下面的迭代式成立：

$$a_k=\frac{1}{6}(1-a_{k-1})+\frac{5}{6}a_{k-1}, \qquad a_0=1 \tag{4.46}$$

(2) 试求 $\{a_k; k=0,1,2,\cdots\}$ 的母函数 $A(z)$.

(3) 试求 a_k.

5. 某公司把 N 种彩券中的一种随机地放到该公司出售的每盒糕点箱中. 如果某人能够收集到 r 种 $(r<N)$ 不同的彩券，就能够得到一份奖品. 有一个小孩，每次买一盒糕点. 我们假定当他买到第 n 盒糕点的时候，第一次收集全 r 种不同彩券的概率为 $p(n;r)$. 很显然，当 $n<r$ 的时候，$p(n;r)=0$.

(1) 当 $n\geq r$ 的时候，试证下面等式成立.

$$\begin{aligned}p(n;r)=&p(n-1;r-1)(1-\frac{r-1}{N})+p(n-2;r-1)\frac{r-1}{N}(1-\frac{r-1}{N})\\&+p(n-3;r-1)(\frac{r-1}{N})^2(1-\frac{r-1}{N})\\&+p(n-4;r-1)(\frac{r-1}{N})^3(1-\frac{r-1}{N})\\&+\cdots+p(r-1;r-1)(\frac{r-1}{N})^{n-r}(1-\frac{r-1}{N})\end{aligned} \tag{4.47}$$

(2) 根据上面结果，试证下面等式成立.

$$p(n+1;r) = \frac{r-1}{N}p(n;r) + (1 - \frac{r-1}{N})p(n;r-1) \tag{4.48}$$

(3) 假定概率分布 $\{p(n;r); n = 0, 1, 2, \cdots\}$ 的母函数为

$$P(z;r) = \sum_{n=0}^{\infty} p(n;r)z^n$$

试证

$$\{N-(r-1)z\}p(z;r) = (N-r+1)zP(z;r-1) \tag{4.49}$$

(4) 试证 $P(z;r)$ 可写成如下形式：

$$P(z;r) = z\prod_{j=1}^{r-1} \frac{(N-j)z}{N-jz} \tag{4.50}$$

(5) 试证这个概率分布 $\{p(n;r); n = 0, 1, 2, \cdots\}$ 相当于 "(r-1) 个相互独立且服从 (不相同的) 几何分布的随机变量之和" $+r$ 的分布.

6. 上一题中，如果某人收集到 r 种该公司指定的不同彩券才能得到一份奖品. 试问上题又该如何解答.

7. 设随机变量 N 只取非负整数值，且其分布的概率母函数为 $G_N(z)$. 又设相互独立的随机变量 $X_1, X_2, \cdots$ 与 N 相互独立地服从同一分布，且其矩母函数为 $M_X(\theta)$. 试证 "不定个数 X 之和" $S_N = X_1 + X_2 + \cdots + X_N (S_0 = 0)$ 的分布的矩母函数 $M_s(0)$ 等于 $G_N(M_X(\theta))$. 另外，根据这个结果，试推出习题 3 问题 8 的结论.

8. 概率密度函数为

$$f(x) = \frac{\alpha}{\pi[(x-\mu)^2 + \alpha^2]} \qquad (-\infty < x < \infty)$$

的柯西分布 $C(\mu,\alpha)$ 的特征函数为 $\varphi(t) = \exp[iut - \alpha|t|]$ (见表 3-1). 试证如果相互独立的 $X_1, X_2, \cdots, X_n$ 服从同一柯西分布, $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$ 也服从与同一柯西分布.

9. 设相互独立的随机变量 $X_1, X_2, \cdots, X_n$ 服从标准正态分布 N(0,1). 试证下面通过正交变换得到的随机变量 $Y_1, Y_2, \cdots, Y_n$ 仍相互独立地服从标准正态分布 N(0,1).[提示：利用特征函数]

$$Y_k = a_{k1}X_1 + A_{k2}X_2 + \cdots + a_{kn}X_n \qquad (1 \le k \le n)$$

$$\sum_{j=1}^{n} a_{kj}a_{lj} = \begin{cases} 0 & (k \neq l) \\ 1 & (k = l) \end{cases}$$

第 5 章　泊松过程

5.1　随机过程的基本概念

下面，我们讨论有关随机过程的话题．随机过程，就好比某一固定场所的气温变化情况一样，简单来讲是指随着时间的推移而变化的量的过程．随机过程论就是对于这种量的变化进行的有关分布及数学期望等概率方面的讨论．首先，让我们举两个简单的例子．

例 5-1(伯努利试验过程)　掷 5 次骰子，每次掷出的点数为偶数的话记为 0, 奇数的话记为 1．我们把这类试验称做 **伯努利试验**．

图 5-1 的 (a) 和 (b) 分别表示某同学 A 和某同学 B 进行的试验结果．这里，横轴表示掷骰子的顺序，纵轴表示掷骰子的结果．图的上方标出了掷出的数字．无论谁做这样的试验，都有可能得到与同学 A 或同学 B 相同的结果，也有可能得到与他们相异的结果．

一般来讲，通过试验或观测收集到的数据伴随着概率变化．或者说，这种以产生数据为背景的过程有概率性的变动．我们把这种过程称做 **随机过程**，把在这种过程当中得到的数据按其发生的顺序进行的排列称做随机过程的 **样本函数** (sample function)．图 5-1 就是一个样本函数的例子．

下面我们把伯努利试验的内容重新整理叙述如下．

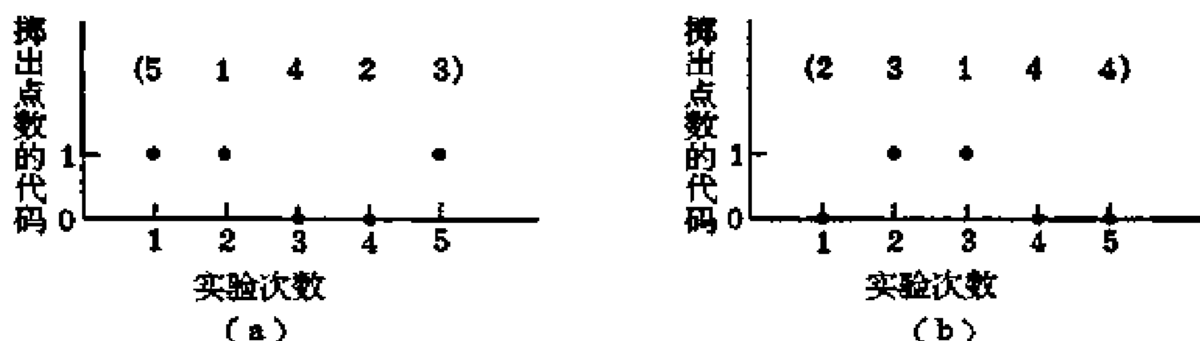


图 5-1 掷骰子的试验结果

掷 5 次骰子得到的所有可能结果共有 6^5 种，其中包括同学 A 掷出的 (5,1,4,2,3) 和同学 B 掷出的 (2,3,1,4,4). 这些所有可能结果的集合被称为 **样本空间** Ω . 即令 $\omega=(w_1,w_2,w_3,w_4,w_5)$, 则有

$$\Omega=\{\omega:\text{对于}n(1\leq n\leq 5),\omega_n=1,2,3,4,5\text{或为}6\}$$

又在 6^5 种可能出现的结果中，每一种结果出现的概率均相等. 如果对于每一个 $\omega\in\Omega$ 和每一个 $n(1\leq n\leq 5)$, 我们定义

$$X_n(\omega)=\begin{cases}0 & (\omega_n\text{为偶数})\\ 1 & (\omega_n\text{为奇数})\end{cases}$$

则 $X_n(\omega)$ 表示掷第 n 次骰子所可能出现结果的随机变量. 这里要注意，随机变量 $X_1(\omega),X_2(\omega),\cdots,X_5(\omega)$ 都是定义在同一样本空间 Ω 上的. 这 5 个随机变量按顺序进行的排列 $\{X_n(\omega);n=1,2,3,4,5\}$ 就是伯努利试验过程. 上面讲了随机过程的一个例子. 从 Ω 中拿出一个 ω, 便可得到如图 5-1 的曲线，即可以确定一个样本函数.

例 5-2(泊松过程) 为了考察能否缓和计算中心的混乱状况，我们通过计算机编程来进行试验. 这类试验被称为计算机模拟.

首先，为了搞清工作(计算请求)到来的情况，我们通过几天的记录发现它大约每分钟发生两件且到来间隔服从指数分布. 于是，为了进行计算机模拟，我们需要得到服从均值为 0.5 分的指数分布的随机数，而这可以简单地通过大多数计算机内所备有的均匀随机数(服从均匀分布的随机数)得到. 我们假定 $\omega = (\omega_1, \omega_2, \omega_3, \cdots)$ 是这样的均匀随机数序列，即对于所有的 $n \geq 1$，都有 $0 < \omega_n < 1$ 且 ω_n 相互独立. 然后令

$$D_n = -0.5 \log \omega_n \quad (\log \text{为自然对数})$$

则 $D_1, D_2, \cdots$ 是服从数学期望为 0.5 分的指数分布且相互独立的随机变量序列[1] 图 5-2 就是按以上述方法得到的一个随机序列模拟的工作到来情况，这里横轴 t 表示经过的时间，纵轴表示在时间 $(0, t]$ 内到来的工作件数的总数. 同理，对于不同的随机序列，我们可以得到不同的曲线.

这个试验的样本空间为

$$\Omega = \{\omega : \omega = (\omega_1, \omega_2, \omega_3, \cdots), \text{对于} n \geq 1, 0 < \omega_n < 1\}$$

对于每个 ω，设 $X_t(\omega)$ 为到时刻 t 为止到来的工作件数的累计值，如果对于

$$-0.5 \sum_{n=1}^{i} \log \omega_n \leq t < -0.5 \sum_{n=1}^{i+1} \log \omega_n$$

有 $X_t(\omega) = i$，则由所有满足 $t \geq 0$ 的 $X_t(w)$ 构成的集合 $\{X_t(\omega); t \geq 0\}$ 就是一个随机过程. 固定一个 t，则 $X_t(\omega)$ 是定义在 Ω 上的随

[1] 设 τ 为正定常数，则 $P\{D_n \leq \tau\} = P\{-0.5 \log \omega_n \leq \tau\} = P\{\omega_n \geq e^{-2\tau}\} = 1 - e^{-2\tau}$ 很显然，D_n 服从均值为 0.5 的指数分布，且它们相互独立.



图 5-2　计算中心的工作到来件数

机变量. 反过来, 固定一个 ω, 则 $X_t(\omega)$ 就变成如图 5-2 所示的 t 的函数, 即样本函数.

正如上面两个例题所述, 一般来讲, 我们把随机过程看成样本函数的形式, 但进行理论展开的时候, 则把它看成随机变量的集合. 随机过程通常表示成 $\{X_t(\omega); t \in T\}$ 的形式. 多数场合下, t 表示时刻或顺序的变量, **指标集合** (index set) T 表示 t 的变化范围. 例 5-1 中 $T = \{1, 2, 3, 4, 5\}$, 例 5-2 中 T 是非负实数全体. 例 5-1 是 **离散时间的随机过程** (process of discrete time), 例 5-2 是 **连续时间的随机过程** (process of continuous time).

如果用 $X_t(\omega)$ 表示时刻 t 时随机过程的 **状态** (state), 那么所有可能出现的 $X_t(\omega)$ (所有 $t \in T$) 的值的集合 S 被称为这个随机过程的 **状态空间**. 例 5-1 中 $S = \{0, 1\}$, 例 5-2 中 S 为非负整数全体. 以上都是 **离散状态空间** (discrete state) 的例子, 而现实生活中, 还存在许多诸如一天气温变化的自动纪录等 **连续状态空间** (continuous state) 的例子.

如上所述, 随机过程的正规表示方法为 $\{X_t(\omega); t \in T\}$, 但在很多参考书中, 大多忽略 ω (或者 Ω) 而单用 $\{X_t; t \in T\}$ 或者

$\{X(t); t \in T\}$ 表示它. 今后，我们也将采取这种表记方法. 特别地，如果 t 的变化范围 T 不致引起混乱，我们还可把它简写成 $\{X(t)\}$ 或者 $X(t)$.

假定常数 $t_0, t_1 (t_0 < t_1)$ 属于 T. 这时，我们称 $X(t_1) - X(t_0)$ 为随机过程 $\{X(t); t \in T\}$ 的从 $t = t_0$ 到 $t = t_1$ 的 **增量** (increment). 对于属于 T 的任意指标组 $t_0 < t_1 < \cdots < t_n$ (n 可取任意值), 如果 n 个增量 $X(t_1) - X(t_0), X(t_2) - X(t_1), \cdots, X(t_n) - X(t_{n-1})$ 相互独立，我们称这个随机过程是 **独立增量的**(stochastic process with independent increments). 此外，对于任意 $s > 0$, 如果两个增量 $X(t_1 + s) - X(t_0 + s)$ 和 $X(t_1) - X(t_0)$ 服从同一分布，我们称这个随机过程是 **定常独立增量的**(stochastic process with stationary independent increments). 刚才讲的两个例子都是定常独立增量的.

5.2 泊松过程的性质

计数过程 (counting process) 是一个理论上比较简单而在实际中常被使用的随机过程. 正如例 5-2 描述的工作到来情况，它记录的是所考察特定现象的发生次数随时间发生的变化. 很显然，一般来讲，它是以非负整数全体为其状态空间的连续时间过程. 从 5-3 我们可以看出，此样本函数是一个单调不减且右连续的阶梯型函数.

下面定义计数过程的有关记号. 设开始观测时 $t = 0$, 所考察现象的发生时刻分别为 $T_1, T_2, T_3, \cdots$, 现象发生的时间间隔为 $D_1(= T_1), D_2(= T_2 - T_1), D_3(= T_3 - T_2), \cdots$. 另外，我们用 $N(t)$ 表示 $(0, t]$ 之间发生的现象的次数， $N(0) = 0$. 泊松过程是理论上比较简单而在实际中常被使用的计数过程. 正如例 5-2 所示，它可

以如下定义.

定义 现象发生的时间间隔 $D_1, D_2, \cdots$ 相互独立且服从同一指数分布的计数过程 $\{N(t); t \geq 0\}$ 称做 **泊松过程**.

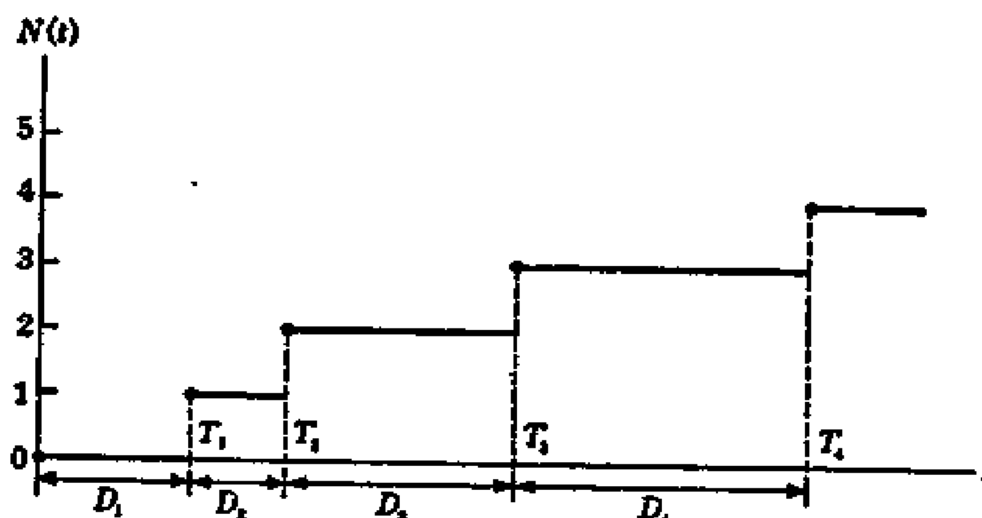


图 5-3 泊松过程的样本函数例

在现象发生的时间间隔服从 $\mathrm{Ex}(\lambda)$ 的假定下，下面的定理给出了这个过程之所以被称为泊松过程的理由.

定理 5-1 对于任意固定的 t, $N(t)$ 服从泊松分布：

$$P(N(t)=n)=e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad (n=0,1,2,\cdots) \tag{5.1}$$

证明 首先，考察 $n=0$ 的情况. 因为 $N(t)=0$ 与 $D_1>t$ 等价，故

$$P\{N(t)=0\}=P\{D_1>t\}=e^{-\lambda t}$$

其次，考察 $n \geq 1$ 的情况．事件 $\{N(t)=n\}$ 与事件 $\{T_n \leq t, T_{n+1} > t\} = \{T_n \leq t, D_{n+1} > t - T_n\}$ 等价．而 $T_n = D_1 + D_2 + \cdots + D_n$ 是相互独立地服从指数分布 Ex(λ) 的随机变量之和，所以根据第二章的知识，我们知道它服从伽玛分布 G(λ,n)，其概率密度函数为

$$f_n(\tau) = \lambda e^{-\lambda t}\frac{(\lambda\tau)^{n-1}}{(n-1)!} \quad (\tau \geq 0) \tag{5.2}$$

于是由全概率公式，我们可以推出式 (5-1)：

$$\begin{aligned} P\{N(t)=n\} &= \int_0^t P\{T_n \leq t, D_{n+1} > t - T_n | T_n = \tau\} f_n(\tau)\mathrm{d}\tau \\ &= \int_0^t P\{D_{n+1} > t - \tau\} f_n(\tau)\mathrm{d}\tau \\ &= \int_0^t e^{-\lambda(t-\tau)} \cdot \lambda e^{-\lambda\tau}\frac{(\lambda\tau)^{n-1}}{(n-1)!}\mathrm{d}\tau \\ &= e^{-\lambda t}\frac{(\lambda t)^n}{n!} \end{aligned}$$

因为泊松分布 (5.1) 的均值为 λt，所以 λ 称做 $(0,t]$ 内单位时间现象发生的**生起率** (intensity)．下面我们考察任意时间间隔 $(t,t+s]$ 的生起率，或者更一般地，考察这个时间间隔的增量 $N(t+s)-N(t)$．

如果 t 与现象发生的时刻一致，这个问题容易解决．因为它相当于利用 s 那么一段时间去观察从时刻 t 开始的泊松过程，即 $N(t+s)-N(t)$ 服从均值为 λs 的泊松分布，且 $N(t+s)-N(t)$ 与 $N(t)$ 相互独立．

如果 t 与现象发生的时刻不一致，**指数分布的马尔可夫性** (Markov property of exponential distribution) 将起到重要作用．

定理 5-2 如果随机变量 D 服从指数分布，那么对于任意 $x, y (\geq 0)$，

$$P\{D > x + y | D > x\} = P\{D > y\} \tag{5.3}$$

证明 由条件概率公式 (式 (1.19)),

$$\begin{aligned}P\{D > x + y | D > x\} &= P\{D > x + y, D > x\}/P\{D > x\} \\ &= P\{D > x + y\}/P\{D > x\} \\ &= e^{-\lambda(x+y)}/e^{-\lambda x} = e^{-\lambda y}\end{aligned}$$

这个性质表明，对于现象的发生间隔服从指数分布的随机过程，从最后一个现象发生开始到现在所经历的时间对于从现在开始到下一个现象发生所经历的时间没有任何影响．这里，我们还可以推出下面的定理．

定理 5-3 泊松过程 $\{N(t); t \geq 0\}$ 是独立增量的，且对于任意实数 $s, t (\geq 0)$, 增量 $N(t+s) - N(t)$ 服从数学期望为 λs 的泊松分布：

$$P\{N(t+s) - N(t) = n\} = e^{-\lambda s}\frac{(\lambda s)^n}{n!}(n = 0, 1, 2, \cdots) \qquad (5.4)$$

反过来，其逆定也成立．

定理 5-4 如果计数过程 $\{N(t); t \geq 0\}$ 为独立增量的且对于任意实数 $s, t (\geq 0)$, 增量 $N(t+s) - N(t)$ 服从数学期望为 λs 的泊松分布 (5.4), 那么这个随机过程服从生起率为 λ 的泊松分布．

我们可以如下理解定理 5-4. 在式 (5.4) 中令 $n = 0$, 则有

$$P\{N(t+s) - N(t) = 0\} = e^{-\lambda s}$$

这表明，从任意时刻 t 开始观测起到下一个现象发生为止所经历的时间服从指数分布．另外又根据增量的独立性，我们容易推知现象的发生间隔 $D_1, D_2, \cdots$ 相互独立地服从同一指数分布．

根据式 (5.4), 我们还可以推出微小时间内现象发生概率的性质.

$$P\{N(t+h)-N(t)=1\}=\lambda h+o(h) \tag{5.5}$$

$$P\{N(t+h)-N(t)\geq 2\}=o(h) \tag{5.6}$$

细心的读者会发现，由式 (5.5),(5.6) 及增量的独立性，我们反过来可以推出式 (5.4).

定理 5-5 如果计数过程 $\{N(t);t\geq 0\}$ 是独立增量的，且对于任意实数 $t(\geq 0)$, 如果式 (5.5) 及式 (5.6) 成立，那么这个随机过程是泊松过程. [1]

例题 5-1 甲、乙两路公共汽车都通过某一车站. 两路公共汽车的到达分别独立地服从 10 分钟 1 辆 (甲), 15 分钟 1 辆 (乙) 的泊松分布. 假定车总不会满员，试问：

(1) 可乘坐甲或乙两路公共汽车的乘客在此车站所需等待时间的概率分布及其均值.

(2) 只可乘坐乙路公共汽车的乘客在此车站等车的时候，恰好有两辆甲路公共汽车通过的概率.

(3) 在 (2) 的情况下，所通过的甲路公共汽车辆数的概率分布.

(1) 反映甲乙两路公共汽车到达情况的泊松分布 $N_1(t)$ 与 $N_2(t)$ 的生起率分别为 $\lambda_1=1/10$, $\lambda_2=1/15$ (单位都是 [1/ 分]). 下面我们证明两路公共汽车的混和到达过程 $N(t)$ 服从生起率为 $\lambda=\lambda_1+\lambda_2$ 的泊松过程. 首先，我们确定定理 5-4 的条件能够被满足. 显然， $N(t)=N_1(t)+N_2(t)$ 是独立增量的. 另外，由

[1] 它的证明，参考习题 5 中问题 5 及其略解.

于增量 $N(t+s)-N(t)$ 是相互独立地服从泊松分布的随机变量 $N_1(t+s)-N_1(t)$ 及 $N_2(t+s)-N_2(t)$ 的和，故由泊松分布的再生性，它服从均值为 λs 的泊松分布. 因此，$N(t)$ 服从生起率为 $\lambda=1/10+1/15=1/6$ 的泊松过程，公共汽车的到达时间间隔服从均值为 6 分的指数分布. 再由指数分布的马尔可夫性，这位乘客的等待时间也服从均值为 6 分的指数分布.

(2) 假定自这位乘客到达车站开始到乙路公共汽车到达车站为止所经过的时间为 s. 那么这期间正好有两辆甲路公共汽车通过的概率为

$$e^{-\lambda_1 s}\frac{(\lambda_1 s)^2}{2!}$$

因为 s 服从参数为 λ_2 的指数分布，故所求概率

$$\begin{aligned}P_2&=\int_0^\infty e^{-\lambda_1 s}\frac{(\lambda_1 s)^2}{2!}\lambda_2 e^{-\lambda_2 s}\mathrm{d}s\\&=\frac{\lambda_1^2\lambda_2}{(\lambda_1+\lambda_2)^3}\int_0^\infty e^{-(\lambda_1+\lambda_2)s}\frac{\{(\lambda_1+\lambda_2)s\}^2}{2!}(\lambda_1+\lambda_2)\mathrm{d}s\end{aligned}$$

很显然，上式右边的被积函数是伽玛分布 $G(\lambda_1+\lambda_2,3)$ 的概率密度函数 (式 (5.2)), 故其积分为 1. 因而

$$P_2=\frac{\lambda_1^2\lambda_2}{(\lambda_1+\lambda_2)^3}=(\frac{1}{10})^2(\frac{1}{15})/(\frac{1}{6})^3=\frac{18}{125}$$

(3) 一般地，设有 n 辆公共汽车通过的概率为 P_n. 参照 (2), 我们有

$$P_n=\frac{\lambda_1^n\lambda_2}{(\lambda_1+\lambda_2)^{n+1}}=(\frac{\lambda_2}{\lambda_1+\lambda_2})(\frac{\lambda_1}{\lambda_1+\lambda_2})^n\quad(n=0,1,2,\cdots)$$

很显然，它服从几何分布且其均值为

$$\sum_{n=0}^{\infty}nP_n=\lambda_1/\lambda_2=1.5(\text{台})$$

例题 5-2 通过日本东名高速公路某一交叉口的车辆数 $N(t)$ 服从平均 1 分种 3 辆的泊松过程. 其中 60% 驶向东京方向, 40% 驶向名古屋方向. 现假定各车的驶向相互独立, 且通过这个交叉口所用的时间也忽略不计. 试证驶向这两方面的车辆数 $N_1(t)$(东京方面), $N_2(t)$(名古屋方面) 均服从泊松过程, 且这两个过程相互独立.

利用定理 5-4. 很显然, $N_1(t)$ 与 $N_2(t)$ 都是定常独立增量的. 下面推出 $N_1(t)$ 与 $N_2(t)$ 相互独立. 因为某一车辆驶向东京方面的概率为 $p(=0.6)$, 驶向名古屋方面的概率为 $q(=0.4)$, 故

$$P\{N_1(t)=n_1,N_2(t)=n_2\}=P\{N(t)=n_1+n_2,N_1(t)=n_1\}$$
$$=P\{N(t)=n_1+n_2\}\cdot P\{N_1(t)=n_1|N(t)=n_1+n_2\} \quad (5.7)$$

上式中 $P\{N_1(t)=n_1|N(t)=n_1+n_2\}$ 为二项分布的概率, 所以它等于

$$\frac{(n_1+n_2)!}{n_1!n_2!}p^{n_1}q^{n_2}$$

以分为 t 的单位, 则式 (5.7) 的概率为

$$e^{-3t}\frac{(3t)^{n_1+n_2}}{(n_1+n_2)!}\cdot\frac{(n_1+n_2)!}{n_1!n_2!}p^{n_1}q^{n_2}=e^{-3pt}\frac{(3pt)^{n_1}}{n_1!}\cdot e^{-3qt}\frac{(3qt)^{n_2}}{n_2!}$$

由此我们可以推出 $N_1(t)$ 与 $N_2(t)$ 相互独立. 分别服从平均每分种生起率为 $3p=1.8$ 辆, $3q=1.2$ 辆的泊松过程.

例 5-3(尤尔过程) 考察初等生物群体繁殖过程模型. 假定每个生物体与其它生物体的繁殖相互独立地服从参数为 λ 的泊松过程, 并假定每个生物体一个一个地生"孩子", 且不发生死亡. 则它被称为尤尔 (Yule) 过程, 是纯生过程 (只出生, 不死亡的过程, pure birth process) 的一个例子.

假定 $N(t)$ 为时刻 t 时此群体的个数, 且

$$p_n(t)=P\{N(t)=n\} \qquad (n=0,1,\cdots) \quad (5.8)$$

令 $N(t) = n$, 我们考察微小时间 $(t, t+T]$ 内群体个数的增加规律. 显然，它相当于 n 个相互独立的泊松过程的组合，根据例 5-1 的结果，我们知道它服从参数为 $n\lambda$ 的泊松过程. 于是，这期间生两个或两个以上“孩子”的概率为 $o(h)$, 在 $t+h$ 时刻的生物群体个数服从

$$\begin{aligned}p_n(t+h) =& P\{N(t) = n, (t, t+h]\text{时不出生}\} \\ &+ P\{N(t) = n-1, (t, t+h]\text{时出生个数} = 1\} + o(h) \\ =& p_n(t)(1 - n\lambda h) + p_{n-1}(t)(n-1)\lambda h + o(h)\end{aligned}$$

在上式两边分别减去 $p_n(t)$, 并除以 h, 然后令 $h \to 0$, 则有

$$\frac{\mathrm{d}}{\mathrm{d}t}p_n(t) = -n\lambda p_n(t) + (n-1)\lambda p_{n-1}(t) \tag{5.9}$$

这是关于 $p_n(t)$ 的微分差分方程. 其初始条件由时刻 $t = 0$ 时生物群体的个数决定. 令 $N(0) = n_0$, 则初始条件变为

$$p_{n_0}(0) = 1; \qquad p_n(0) = 0. \quad (n \neq n_0) \tag{5.10}$$

如果 $n < n_0$, 则对于任意 $t(\geq 0)$, $p_n(t) = 0$.

如果 $n = n_0$, 则据式 (5.9) 及式 (5.10),

$$\frac{\mathrm{d}}{\mathrm{d}h}p_{n_0}(t) = -n_0\lambda p_{n_0}(t), \quad p_{n_0}(0) = 1$$

这个常微分方程的解为

$$p_{n_0}(t) = e^{-n_0\lambda t} \tag{5.11}$$

如果 $n \geq n_0 + 1$, 假定 $p_{n-1}(t)$ 为已知函数，并以初始条件 $p_n(0) = 0$ 解式 (5.9), 则

$$p_n(t) = \int_0^t e^{-n\lambda(t-s)}(n-1)\lambda p_{n-1}(s)\mathrm{d}s \tag{5.12}$$

最后由这个递推式及式 (5.11), 我们可依次求出 $n = n_0+1, n = n_0+2, \cdots$ 的 $p_n(t)$. 即

$$p_n(t) = \frac{(n-1)!}{(n-n_0)!(n_0-1)!}(e^{-\lambda t})^{n_0}(1-e^{-\lambda t})^{n-n_0} \quad (n \geq n_0) \tag{5.13}$$

这是负二项分布的概率分布，其均值和方差分别为

$$E[N(t)] = n_0 e^{\lambda t}, \quad \mathrm{Var}[N(t)] = n_0 e^{\lambda t}(e^{\lambda t}-1) \tag{5.14}$$

问题 根据数学归纳法，试证式 (5.13) 为式 (5.12) 的解.

5.3 非齐次泊松过程

考察电话交换台每小时的电话记录，我们会发现不同的时间里打来的电话数也不相同. 即电话呼叫 (call) 的生起率随时间发生变化. 这种现象，我们可以通过生起率随时间发生变化的非齐次泊松过程来进行分析.

定义 如果计数过程 $\{N(t); t \geq 0\}$ 是独立增量的，且对于正数 h 有以下两式成立，则称这个过程为 **非齐次泊松过程** (non-homogeneous poisson process).

$$P\{N(t+h)-N(t)=1\} = \lambda(t)h + o(h) \tag{5.15}$$

$$P\{N(t+h)-N(t) \geq 2\} = o(h) \tag{5.16}$$

对于非齐次泊松过程，我们感兴趣的是时间 $(0,t]$ 内现象的发生次数 $N(t)$, 更一般地讲，是 $(t, t+s]$ 内现象发生次数 $N(t+s)-N(t)$ 的概率分布.

定理 5-6 如果

$$m(t) = \int_0^t \lambda(s)\mathrm{d}s \tag{5.17}$$

则 $N(t)$(或者 $N(t+s)-N(t)$) 服从数学期望为 $m(t)$(或者 $m(t+s)-m(t)$) 的泊松分布.

我们称 $m(t)$ 为这个函数的 **均值函数**(value function). 直观地, 可以如下理解这个定理. 假想有这么一个"时钟", 当 $\lambda(t)$ 取值大的时候它走得快, 当 $\lambda(t)$ 取值小的时候它走得慢. 调好这个"时钟", 使得当我们以这个"时钟"为基准去确定生起率的时候, 生起率能够保持恒定. 以这个"时钟"去观测的随机过程属于泊松过程, 且在一定时间内现象的发生次数服从泊松分布. 更具体地讲, 当一般的时钟指向时刻 t 的时候, 我们可认为假想的"时钟"指向时刻 $\tau(t)$, 并定义

$$\tau(t)=m(t) \tag{5.18}$$

(如图 5-4). 那么, 我们可以定义一个新的计数过程, 即

$$M(\tau)=N[m^{-1}(\tau)]$$

[为了便于理解, 我们假定 $\lambda(t)$ 只取正值以保证其能够唯一确定 $m^{-1}(\tau)$]. 很显然, $M(\tau)$ 是独立增量的, 故若令

$$m^{-1}(\tau)=t,\quad m^{-1}(\tau+h')=t+h$$

则由

$$h'=m(t+h)-m(t)=\int_t^{t+h}\lambda(s)\mathrm{d}s=\lambda(t)h+o(h)$$

$$\lim_{h'\to+0}\frac{P\{M(\tau+h')-M(\tau)=1\}}{h'}=\lim_{h\to+0}\frac{P\{N(t+h)-N(t)=1\}}{\lambda(t)h+o(h)}$$

$=1$, 即

$$P\{M(\tau+h')-M(\tau)=1\}=h'+o(h')$$

同理，

$$P\{M(\tau+h')-M(\tau)\geq 2\}=o(h')$$

即我们可以推出 $\{M(\tau);\tau\geq 0\}$ 服从生起率为 1 的泊松过程.

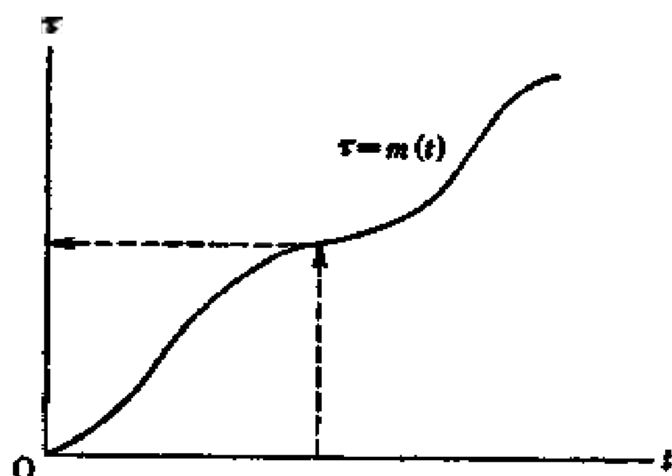


图 5-4 从非齐次泊松过程转为齐次泊松过程时时间变量的转换过程

例题 5-3(设备的故障率) 我们假定某一设备发生故障的次数服从非齐次泊松过程. 图 5-5 给出了自购入这个设备 t 个月以后的故障率 $\lambda(t)$ [次 / 月]. 这个设备的购买费用为 K=21[万元], 修理费用为 C=2[万元 / 回]. 不考虑利息及经济变动, 问时隔几个月更新这个设备是最合适的.

假定时隔 t 个月更新这个设备，那么每月所需费用 (= 损耗费用 + 修理费用) 的数学期望为

$$c(t)=\frac{K+C_m(t)}{t}$$

令 $\dfrac{\mathrm{d}c(t)}{\mathrm{d}t}=0$, 则由 $\dfrac{\mathrm{d}m(t)}{\mathrm{d}t}=\lambda(t)$

$$\lambda(t)t-m(t)=K/C$$

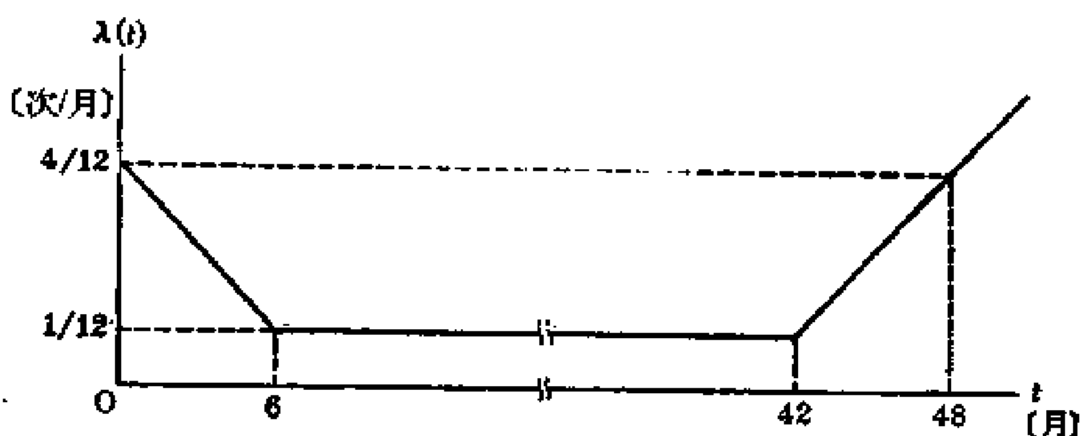


图 5-5　设备的故障率 $\lambda(t)$

令此式左边为 $g(t)$, 则由图 5-5 知 $t \le 42$ 的时候, $g(42) = -3/4$. 而 $t > 42$ 的时候, $\dfrac{\mathrm{d}g}{\mathrm{d}t} = \dfrac{\mathrm{d}\lambda}{\mathrm{d}t}t = t/24 > 0$ 单调递增, 所以根据

$$-3/4 + \int_{42}^{t} \frac{s}{24}\mathrm{d}s = -3/4 + \frac{t^2 - 42^2}{48} = K/C$$

最适值 $t = 48$. 由 $m(48) = 5.5$, 我们可知这个设备在 48 个月内平均发生 5.5 回故障, 耗费修理费用 11 万元, 即大约需要耗费更新费用的一半.

5.4 复合泊松过程

对于泊松过程 $\{N(t); t \ge 0\}$, 当 $N(t)$ 的值在某一时刻发生变化的时候, 其变化量大于等于 2 的概率为 0(式 (5.6)). 这一节, 我们放宽这个条件而考虑变化量取不同值时的情况 (不只局限于整数, 而且也可为负). 很显然, 这个过程不再是计数过程了.

定义　假定 $\{N(t); t \ge 0\}$ 服从泊松过程, 且 $\{Y_n; n = 1, 2, \cdots\}$ 是相互独立地服从同一分布的随机变量序列. 当 $\{N(t)\}$ 与 $\{Y_n\}$

相互独立的时候，定义

$$X(t) = \sum_{n=1}^{N(t)} Y_n \tag{5.19}$$

那么随机过程 $\{X(t); t \geq 0\}$ 称做**复合泊松过程**(compound Poisson process).

图 5-6 描述了复合泊松分布样本函数的一个例子.

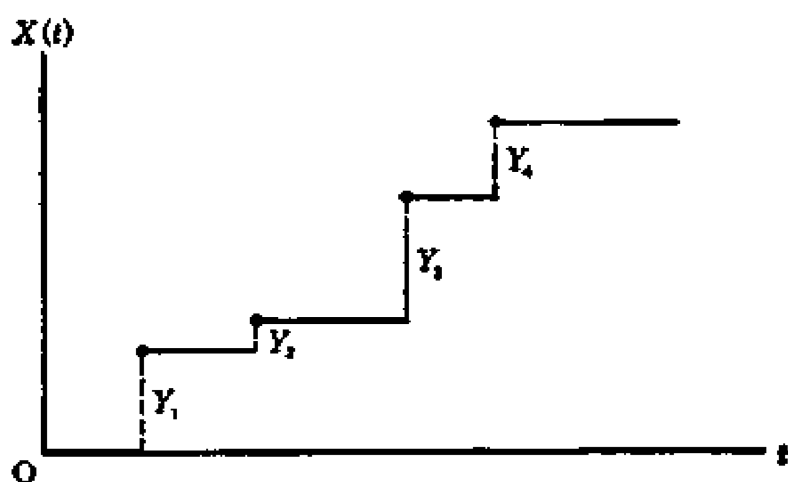


图 5-6 复合泊松过程的样本函数例

比如假定财务科的帐单到来服从泊松分布，处理每个帐单所需时间相互独立地服从同一分布，然后假定 $X(t)$ 为处理到时刻 t 为止到来帐单所需的延迟时间，那么 $\{X(t); t \geq 0\}$ 服从复合泊松分布过程.

$X(t)$ 的分布一般来讲比较复杂. 但 $X(t)$ 的特征函数可根据 Y_n 的特征函数简单地写成如下形式. 即如果令

$$\phi_{X(t)}(u) = E[e^{iuX(t)}], \qquad \phi_{Y(u)} = E[e^{iuY_n}] \tag{5.20}$$

则

$$\phi_{X(t)}(u) = \exp[\lambda t\{\varphi_Y(u) - 1\}] \tag{5.21}$$

根据 $N(t)$ 的条件期望我们可以如下证出上式.

$$E[e^{\mathrm{i}uX(t)}] = \sum_{n=0}^{\infty} E[e^{\mathrm{i}uX(t)}|N(t) = n]e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

这里,

$$\begin{aligned}E[e^{\mathrm{i}uX(t)}|N(t) = n] &= E[e^{\mathrm{i}u(Y_1+Y_2+\cdots+Y_n)}] \\ &= \{E[e^{\mathrm{i}uY_1}]\}^n \\ &= \{\varphi_Y(u)\}^n\end{aligned}$$

于是,

$$\begin{aligned}\varphi_{X(t)}(u) &= \sum_{n=0}^{\infty}\{\phi_y(u)\}^n e^{-\lambda t}\frac{(\lambda t)^n}{n!} = e^{-\lambda t}\cdot\exp[\lambda t\varphi_Y(u)] \\ &= \exp[\lambda t\{\varphi_Y(u) - 1\}]\end{aligned}$$

对上式求导, 再由第 4 章第 3 节的式 (4.26) 及 (4.27),

$$E[X(t)] = \varphi'_{X(t)}(0)/\mathrm{i} = \lambda t E[Y] \tag{5.22}$$

$$\mathrm{Var}[X(t)] = -\phi''_{X(t)}(0) - \{E[X(t)]\}^2 = \lambda t E[Y^2] \tag{5.23}$$

直观地, 我们可以如下理解式 (5.22), 即

$$E[X(t)] = E[N(t)]E[Y] = \lambda t E[Y] \tag{5.24}$$

很显然, 式 (5.23) 的最后一项并不等于 $\lambda t\mathrm{Var}[Y]$. 一般地, 如果 $X = Y_1 + Y_2 + \cdots + Y_n$ (定数个独立同分布的随机变量之和), 那么

$$\mathrm{Var}[X] = n\mathrm{Var}[Y] \tag{5.25}$$

如果 $X' = Y_1 + Y_2 + \cdots + Y_N$ (不定个数独立同分布的随机变量之和), 那么

$$\mathrm{Var}[X'] = E[N]\mathrm{Var}[Y] + \mathrm{Var}[N]\{E[Y]\}^2 \tag{5.26}$$

问题 如果N服从数学期望为 λt 的泊松分布, 试证从式 (5.26), 我们可以推出

$$\mathrm{Var}[X'] = \lambda t E[Y^2]$$

习 题 5

1. 设 $\{N(t); t \geq 0\}$ 服从生起率为 λ 的泊松过程. 试计算下式.

(1) $P\{N(5) = 4\}$

(2) $P\{N(5) = 4, N(7.5) = 6, N(12) = 9\}$

(3) $P\{N(12) = 9|N(5) = 4\}$

(4) $E[N(5)]$

(5) $\mathrm{Var}[N(5)]$

(6) $E[N(10)|N(5) = 4]$

2. 某一公共汽车站的乘客到来服从平均1分种1人的泊松过程, 公共汽车的运行间隔服从8分至12分的均匀分布. 试求自某一公共汽车开出此站到另一公共汽车驶入此站所到乘客数的数学期望和方差.

3. 试求服从生起率为 λ 的泊松过程 $N(t)$ 的协方差 $\mathrm{Cov}(N(t), N(t+s))$.

4. 某一加油站的车辆到来服从平均2分种1辆的泊松过程. 假定各车的加油量为 $Y[l]$, $Z = (Y - 10)/40$ 服从贝塔分布 $\mathrm{Be}(3,2)$. 试求一小时内驶入此加油站车辆加油总量的数学期望及方差.

5. 试证定理 5-5. [提示：很显然，只需证明 $t=0$ 的情况. 仿例 5-3, 求出满足 $p_n(s)=P\{N(s)=n\}$ 的微分差分方程，并由此推出 $N(s)$ 服从泊松分布.]

6. 某一车站售票处顾客的到来服从平均 1 分钟 λ 人的泊松过程. 每个顾客买票所需时间服从平均 $1/\mu$ [分] 的指数分布. 我们用 $p_n(t)$ 表示在时刻 t 的时候，售票口前有 n 个顾客的概率.

(1) 对于 $n \geq 1$, 试推出以下关系式：

$$
\begin{aligned}
p_n(t+h)= & p_n(t)(1-\lambda h)(1-\mu h)+p_{n+1}(t)\mu h(1-\lambda h) \\
& +p_{n-1}(t)\lambda h(1-\mu h)+o(h)
\end{aligned}
$$

(2) 当 $n=0$ 的时候，应如何修改上式.

(3) 试推出关于 $p_n(t)$. $(n=0,1,2,\cdots)$ 的微分差分方程.

(4) 经过很长时间以后，如果对于所有 n, 都有 $\dfrac{\mathrm{d}p_n(t)}{\mathrm{d}t}=0$, 我们称它达到了稳态. 这时，我们把 $p_n(t)$ 写成 p_n 的形式. 试求有关 $p_n(n=0,1,2,\cdots)$ 的差分方程.

(5) 解出上面的差分方程. 为了使 $p_n(n=0,1,2,\cdots)$ 呈概率分布 $\sum\limits_{n=0}^{\infty} P_n=1$, 试问 λ 与 μ 应满足什么关系.

第 6 章 再生过程

第 5 章所述泊松过程, 是现象发生的时间间隔相互独立地服从同一指数分布的计数过程. 这一章, 我们讨论现象发生的时间间隔不只局限于指数分布的计数过程. 这个过程称做 **再生过程** (renewal process).[1] 根据不同的场合, 再生有时也称做 **更新**(renewal). 考虑到设备发生故障便更新这一过程, 我们就可以理解用它命名的理由了. 也就是说, 只要进行再生或更新, 过程将从某一新的起点开始出发.

6.1 基本概念

我们用 $T_0 = 0 < T_1 < T_2 < \cdots$ 表示某一特定现象发生的时刻, 并假定此现象发生的时间间隔 $D_1(= T_1 - T_0), D_2(= T_2 - T_1), \cdots$ 相互独立地服从同一分布. 如果这个分布的分布函数为 F, 数学期望为 μ, 方差为 σ^2 [2] 现在, 用 $N(t)$ 来表示在 $(0, t]$ 内现象发生的次数:

$$N(t) = \max\{n : T_n \leq t\} \tag{6.1}$$

我们称此随机过程 $\{N(t); t \geq 0\}$ 为 **再生过程**.

首先, 考察 $N(t)$ 的分布与 F 之间的关系. 从式 (6.1)(或直观地), 我们容易得到如下的等价关系.

$$N(t) \geq n \Leftrightarrow T_n \leq t \tag{6.2}$$

---

[1] 再生过程有时也指现象发生的 "时刻" 而不是现象发生的 "次数".

[2] 本书令 $F(0) = 0$, $F(\infty) = 1$, 即现象发生的时间间隔必为有限的正数. 其它场合, 我们也可进行类似讨论.

另一方面，由

$$T_n = D_1 + D_2 + \cdots + D_n \tag{6.3}$$

我们得到第 2 章第 5 节所述

$$P\{T_n \le t\} = F^{(n)}(t) \tag{6.4}$$

这里， $F^{(n)}$ 为 F 的 n 重卷积， $F^{(0)}(t) = 1$. 于是，

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \ge n\} - P\{N(t) \ge n+1\} \\ &= F^{(n)}(t) - F^{(n+1)}(t) \end{aligned} \tag{6.5}$$

我们用 $m(t)$ 来表示 $N(t)$ 的均值：

$$m(t) = E[N(t)] \tag{6.6}$$

$m(t)$ 称做 **均值函数**, 或 **再生函数** (renewal function).

定理 6-1

$$m(t) = \sum_{n=1}^{\infty} F^{(n)}(t) \tag{6.7}$$

证明

$$\begin{aligned} \sum_{n=1}^{\infty} F^{(n)}(t) &= \sum_{n=1}^{\infty} P\{T_n \le t\} \\ &= \sum_{n=1}^{\infty} P\{N(t) \ge n\} \\ &= E[N(t)] \end{aligned}$$

这里，最后的等式来自定理 3-2 的推论.

问题 对于泊松过程，试求 $F^{(n)}(t)$ 及 $m(t)$.

6.2 再生方程式

理论上，均值函数 $m(t)$ 可以通过式 (6.7) 求出，而实际上这不容易做到. 这一节，我们推出 $m(t)$ 的另一形式 (积分方程式). 它对于考察 $t\to\infty$ 时 $m(t)$ 的变化情况非常方便.

根据 T_1 的条件期望，我们可将 $m(t)=E[N(t)]$ 写成如下形式：

$$m(t)=\int_0^\infty E[N(t)|T_1=s]\mathrm{d}F(s) \tag{6.8}$$

当 $T_1=s>t$ 的时候，N(t)=0. 当 $T_1=s\le t$ 的时候，$E[N(t)|T_1=s]=1+E[N(t-s)]$, 故式 (6.8) 变为

$$\begin{aligned}m(t)=&\int_0^t\{1+m(t-s)\}\mathrm{d}F(s)\\=&F(t)+\int_0^t m(t-s)\mathrm{d}F(s)\end{aligned} \tag{6.9}$$

这就是 **再生方程式** (renewal equation).

将式 (6.9) 一般化得到积分方程式

$$g(t)=h(t)+\int_0^t g(t-s)\mathrm{d}F(s)\quad (t\ge 0) \tag{6.10}$$

它被称做 **再生型方程式** (renewal type equation). 其中， h 和 F 为已知函数， g 为未知函数. g 可以通过拉普拉斯变换 (Laplace transformation) 求出. 即如果定义 [1] $\widetilde{g}(z)=$

$$\int_0^\infty e^{-zt}g(t)\mathrm{d}t,\ \widetilde{h}(z)=\int_0^\infty e^{-zt}h(t)\mathrm{d}t,\ \widetilde{f}(z)=\int_0^\infty e^{-zt}f(t)\mathrm{d}t \tag{6.11}$$

[1] 这里，简单假定 $F(t)$ 服从连续型分布，其概率密度函数为 $f(t)$, 则显然有 $\mathrm{d}F(s)=f(s)\mathrm{d}s$. 对于离散分布，参考习题 6 中问题 8 及其略解.

并在式 (6.10) 两边取拉普拉斯变换，则

$$\tilde{g}(z) = \tilde{h}(z) + \tilde{g}(z)\tilde{f}(z)$$

于是，

$$\tilde{g}(z) = \frac{\tilde{h}(z)}{1-\tilde{f}(z)} \tag{6.12}$$

特别地，对于再生方程式 (6.9)

$$g(t) = m(t), \quad h(t) = F(t) = \int_0^t f(s)\mathrm{d}s$$

又由 $\tilde{F}(z) = \tilde{f}(z)/z$, 我们得到

$$\tilde{m}(z) = \frac{\tilde{f}(z)/z}{1-\tilde{f}(z)} \tag{6.13}$$

最后，对 $\tilde{g}(z)$, $\tilde{m}(z)$ 实行逆变换，即可求出 $g(t)$ 和 $m(t)$.

例 6-1 令 $F(t) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} \quad (t \geq 0)$, 则它是服从均值为 $\mu = 2/\lambda$ 的伽玛分布 $G(\lambda, 2)$ 的分布函数. 虽然，我们可以通过定义式 (6.11) 求出 $\tilde{f}(z)$, 但又由于这个分布相当于相互独立地服从同一指数分布的两个随机变量之和的分布 (第 2 章第 5 节例 2-1), 我们可以简单地推出

$$\tilde{f}(z) = (\frac{\lambda}{\lambda+z})^2$$

根据式 (6.13),

$$\tilde{m}(z) = \frac{\tilde{f}(z)/z}{1-\tilde{f}(z)} = \frac{\lambda^2}{z^2(z+2\lambda)} = \frac{\lambda}{2z^2} - \frac{1}{4}(\frac{1}{z} - \frac{1}{z+2\lambda})$$

再通过变换，最后得到

$$m(t) = \frac{\lambda}{2}t - \frac{1}{4}(1 - e^{-2\lambda t}) = \frac{t}{\mu} - \frac{1}{4} + \frac{1}{4}e^{-2\lambda t} \tag{6.14}$$

例 6-2 (余命分布)(excess life distribution) 我们把从开始观测时刻 t 到下次再生所经历的再生过程的时间 $W(t) = T_{N(t)+1} - t$ 称做 t 的余命. 比如假定公共汽车的到来服从再生过程, 那么余命 $W(t)$ 实际上就是指在时刻 t 来到公共汽车站的乘客所需的等待时间.

为了求出 $W(t)$ 的分布函数, 令 x 为正定常数, 并且

$$Q_x(t) = P\{W(t) > x\} \tag{6.15}$$

则根据最初再生时刻 T_1 的条件概率, 我们可以求出

$$Q_x(t) = \int_0^{\infty} P\{W(t) > x | T_1 = s\} \mathrm{d}F(s) \tag{6.16}$$

而随机过程从 $T_1 = s$ 开始重新出发, 即

$$P\{W(t) > x | T_1 = s\} = \begin{cases} Q_x(t-s) & (s \le t) \\ 0 & (t < s \le t+x) \\ 1 & (t+x < s) \end{cases}$$

因而, 式 (6.16) 将变为

$$\begin{aligned} Q_x(t) &= \int_0^t Q_x(t-s)\mathrm{d}F(s) + \int_{t+x}^{\infty} \mathrm{d}F(s) \\ &= 1 - F(t+x) + \int_0^t Q_x(t-s)\mathrm{d}F(s) \end{aligned} \tag{6.17}$$

它就是决定 $Q_x(t) = P\{W(t) > x\}$ 的再生型方程式. 比如当 $F(t) = 1 - e^{-\lambda t}$ (泊松过程) 的时候, 式 (6.17) 变为

$$Q_x(t) = e^{-\lambda x} \cdot e^{-\lambda t} + \int_0^t Q_x(t-s)\mathrm{d}F(s)$$

对上式两边取拉普拉斯变换, 则

$$\widetilde{Q}_x(z) = e^{-\lambda x}\frac{1}{z+\lambda} + \widetilde{Q}_x(z)\frac{\lambda}{z+\lambda}$$

解之，得到

$$\widetilde{Q}_x(z) = e^{-\lambda x} \cdot \frac{1}{z}$$

最后通过逆变换，我们得到

$$Q_x(t) = e^{-\lambda x}$$

正如第 5 章所述，对于泊松分布，其余命分布并不依赖于观测开始时刻 t.

6.3 极限定理

根据式 (6.7) 或式 (6.9), 我们可以考察 $t \to \infty$ 时 $m(t)$ 的变化情况. 根据拉普拉斯变换的有关理论，$t \to \infty$ 时 $m(t)$ 的变化情况，可通过 $z \to 0$ 时 $\widetilde{m}(z)$ 的变化情况推知. 因为在 $z = 0$ 附近对 e^{-zt} 进行泰勒展开，则

$$e^{-zt} = 1 - zt + \frac{(zt)^2}{2} + O(z^3)$$

故

$$\begin{aligned}\widetilde{f}(z) &= \int_0^\infty e^{-zt} f(t)\mathrm{d}t = 1 - z\int_0^\infty tf(t)\mathrm{d}t + \frac{z^2}{2}\int_0^\infty t^2 f(t)\mathrm{d}t + \cdots \\ &= 1 - z\mu + \frac{z^2}{2}(\mu^2 + \delta^2) + O(z^3)\end{aligned}$$

把上式代入式 (6.13), 并在 $z = 0$ 附近进行泰勒展开，则

$$\widetilde{m}(z) = \frac{1}{\mu z^2} + \frac{\delta^2 - \mu^2}{2\mu^2 z} + O(1)$$

再通过逆变换，

$$m(t) = \frac{t}{\mu} + \frac{\delta^2 - \mu^2}{2\mu^2} + o(1) \quad (t \to \infty) \tag{6.18}$$

因而

$$\lim_{t\to\infty}\frac{m(t)}{t}=\frac{1}{\mu} \tag{6.19}$$

此外，对于 $N(t)$, 我们还知道

$$P\{\lim_{t\to\infty}\frac{N(t)}{t}=\frac{1}{\mu}\}=1 \tag{6.20}$$

式 (6.19) 及 (6.20) 是再生过程论中非常重要的两个基本定理. 严格的证明比较烦琐，但直观地看它们很好理解.

从式 (6.18), 我们还会想到可能对于任意 $a>0$,

$$\lim_{t\to\infty}\{m(t+a)-m(t)\}=\frac{a}{\mu}$$

而事实上，虽然上式在 D 服从连续分布的时候成立，存在服从离散分布的时候未必成立. 下面的定理详细叙述了这部分内容.

定理 6-2　布拉克威尔定理(Blackwell theorem)

(1) 当 D 不服从格状分布 [1] (见第 2 章第 2 节) 的时候，对于任意 $a>0$, 有

$$\lim_{t\to\infty}\{m(t+a)-m(t)\}=\frac{a}{\mu} \tag{6.21}$$

(2) 当 D 服从格子状分布的时候，

$$\lim_{t\to\infty}P\{t=nd\ \ \text{时再生}\}=\frac{d}{\mu} \tag{6.22}$$

这个定理推广到一般情况，我们得到如下定理.

定理 6-3　再生理论的主要定理 (key renewal theorem)

[1] 正如第 2 章第 2 节所述，服从格状分布是指存在 $d>0$, 使得 $\sum_{n=1}^{\infty}P\{D=nd\}=1$ 成立. 满足条件的最大 d 称做此格状分布的 **周期** (period of a lattice distribution).

F 不服从格子状分布，且函数 $h(t) \geq 0$ 满足下面任一条件，则有

$$\lim_{t\to\infty}\int_0^t h(t-s)\mathrm{d}m(s) = \frac{1}{\mu}\int_0^\infty h(s)\mathrm{d}s \tag{6.23}$$

成立.

(1) $h(t)$ 连续且当 t 大于某定值的时候，$h(t) \equiv 0$.

(2) $h(t)$ 单调不增且 $\int_0^\infty h(t)\mathrm{d}t < \infty$.

这个定理，在更一般的条件 ($h(t)$ 为黎曼可积条件) 下也能成立，这里我们不做深入考察.

例 6-3 试求 $t \to \infty$ 时再生型方程式 (6.10) 的解 $g(t)$ 的极限值.

由对式 (6.10) 进行拉普拉斯变换得到的式 (6.12) 及式 (6.13)，有

$$\tilde{g}(z) = \frac{\tilde{f}(z)}{1-\tilde{f}(z)} = \tilde{h}(z) + \tilde{h}(z)\cdot z\,\tilde{m}(z)$$

上式通过逆变换，又有

$$g(t) = h(t) + \int_0^t h(t-s)\mathrm{d}m(s) \tag{6.24}$$

所以，当 F 和 h 满足主要定理条件的时候，

$$\lim_{t\to\infty} g(t) = \lim_{t\to\infty}\int_0^t h(t-s)\mathrm{d}m(s) = \frac{1}{\mu}\int_0^\infty h(s)\mathrm{d}s \tag{6.25}$$

根据这一结果，我们能够求出例 6-2 中 $t \to \infty$ 时余命分布的形

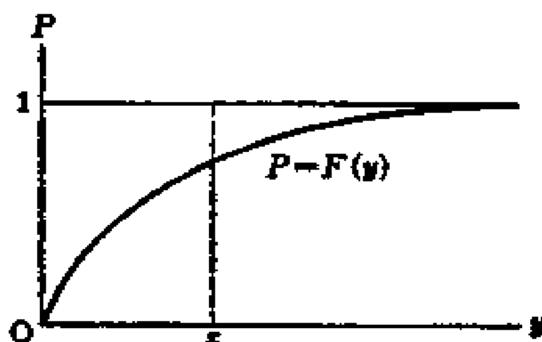
状. 即如果寿命分布 F 不服从格状分布, 由式 (6.17),

$$\lim_{t\to\infty} P\{W(t) > x\} = \frac{1}{\mu}\int_0^{\infty}\{1-F(x+s)\}\mathrm{d}s$$

$$= \frac{1}{\mu}\int_x^{\infty}\{1-F(y)\}\mathrm{d}y$$

$$\lim_{t\to\infty} P\{W(t) \le x\} = \frac{1}{\mu}\int_0^{x}\{1-F(y)\}\mathrm{d}y \tag{6.26}$$

问题 式 (6.26) 的右边相当于图中两部分的面积之比, 试问是哪两部分的面积之比.



例 6-4 (设备的工作率) 某一设备经常发生故障, 且发生故障就马上通过修理使它正常运转. 修理所需时间 X 服从分布 F, 设备自修好到下次发生故障所需时间 Y 服从分布 G, 且它们相互独立. 试问这台设备的工作率.

假定在 $t=0$ 的时候机器开始运转. 首先求在时刻 t 时这台设备正在工作的概率. 由于这个随机过程在修理结束的时候开始新的再生, 令 H 为 $T=X+Y$ 的分布函数, 则根据 T_1 的条件概率我们得到

$$A(t) = \int_0^{\infty} P\{t\text{时运转}|T_1 = s\}\mathrm{d}H(s)$$

$$P\{t\text{时运转}|T_1 = s\} = \begin{cases} A(t-s) & (s \le t) \\ P\{Y_1 > t|T_1 = s\} & (s > t) \end{cases}$$

又由于

$$\int_t^{\infty} P\{Y_1 > t | T_1 = s\} \mathrm{d}H(s) = P\{Y_1 > t, X_1 + Y_1 > t\}$$
$$= P\{Y_1 > t\} = 1 - G(t)$$

我们得到关于 $A(t)$ 的再生型方程式

$$A(t) = 1 - G(t) + \int_0^t A(t-s) \mathrm{d}H(s) \tag{6.27}$$

如果 $T = X + Y$ 不服从格状分布的时候,

$$\lim_{t \to \infty} A(t) = \frac{1}{E[T]} \int_0^{\infty} \{1 - G(t)\} \mathrm{d}t = \frac{E[Y]}{E[X] + E[Y]} \tag{6.28}$$

在信赖性理论中，我们把 $A(t)$ 称做 **有效性** (availability), 把 $A(\infty)$ 称做 **定常有效性**.

例 6-5 (分支过程)(branching process) 某一生物体的寿命分布函数为 F, 且其生命终了时留下 j 个 “孩子” 的概率为 $p_j (j = 0, 1, 2, \cdots)$. 假定每个生物体的寿命及留下的孩子数彼此相互独立. 用 $N(t)$ 表示时刻 t 时生存着的生物体个数，那么随机过程 $\{N(t); t \geq 0\}$ 称做 **分支过程** (branching process). 图 6-1 给出了生物体的个数发生变化的一个例子. 从这个图我们可以看出，它之所以用 “分支” 来命名的理由.

很显然, 如果每个生物体留下孩子数的均值 $a = \sum_{j=0}^{\infty} jp_j$ 小于等于 1, 这一生物体将灭种. 反过来, 如果 a 大于 1, 这一生物体的个数将以正的概率无限增加. 下面就后面的场合，考察 $m(t) = E[N(t)]$ 的变化情况. 这里，我们假定 $t = 0$ 的时候出生的生物体中只有一个能够生存，且 F 不服从格状分布.

假定最初生物体寿命为 T_1, 则有

$$m(t) = \int_0^{\infty} E[N(t) | T_1 = s] \mathrm{d}F(s)$$

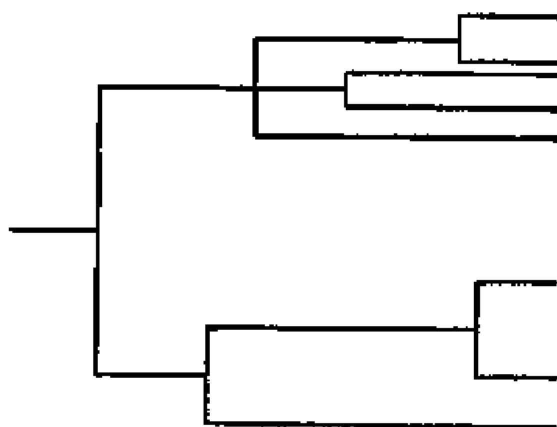


图 6-1 分支过程

又由

$$E[N(t)|T_1 = s] = \begin{cases} a \cdot m(t-s) & (s \leq t) \\ 1 & (s > t) \end{cases}$$

$$m(t) = 1 - F(t) + a\int_0^t m(t-s)\mathrm{d}F(s) \tag{6.29}$$

上式类同于再生型方程式，但由于 a 不等于 1, 因而并非与它完全一致. 下面，通过适当的变换，我们试着导出再生型方程式.

曾在第 5 章讨论过的尤尔过程可以看做是分支过程的特殊情况. 即它是寿命分布 F 服从指数分布且每个生物体在生出两个“孩子”后必然死亡的例子. 在尤尔分布中， $m(t)$ 是 t 的指数函数. 在分支过程中，我们同样可以想象到对于适当的 $\beta > 0, t \to \infty$ 时 $g(t) = e^{-\beta t}m(t)$ 将收敛于某一极限. 从而导出 $g(t)$ 的积分方程式. 很显然，在式 (6.29) 的两边乘以 $e^{-\beta t}$, 我们得到

$$g(t) = e^{-\beta t}\{1 - F(t)\} + a\int_0^t e^{-\beta s}g(t-s)\mathrm{d}F(s) \tag{6.30}$$

为了使上式变成再生型方程式，右边第二项需写成如下形式

$$\int_0^t g(t-s)\mathrm{d}F_\beta(s)$$

这里，

$$\mathrm{d}F_\beta(s) = ae^{-\beta s}\mathrm{d}F(s)$$

又由于 $F_\beta(s)$ 应为分布函数，故由

$$F_\beta(s) = \int_0^s ae^{-\beta y}\mathrm{d}F(y) \tag{6.31}$$

及 $F_\beta(\infty)=1$,

$$\int_0^\infty e^{-\beta y}\mathrm{d}F(y) = \frac{1}{a} \tag{6.32}$$

式 (6.32) 的左边是关于 β 的单调递减连续函数，且 $\beta=0$ 时其值为 1, $\beta\to\infty$ 时其值收敛于 1, 故存在一个能够满足式 (6.32) 的 β. 取定这样一个 β, 式 (6.30) 将变为再生型方程式，根据主要定理，$t\to\infty$ 时

$$g(t)\to\int_0^\infty e^{-\beta y}\{1-f(y)\}\mathrm{d}y\Big/\int_0^\infty s\mathrm{d}F_\beta(s) \tag{6.33}$$

很显然，由分部积分及式 (6.32) 可知式 (6.33) 的分子为

$$-\{1-F(y)\}\frac{e^{-\beta y}}{\beta}\Big|_{y=0}^{\infty} - \int_0^\infty \frac{e^{-\beta y}}{\beta}\mathrm{d}F(y) = \frac{1}{\beta}-\frac{1}{\alpha\beta}$$

分母则为

$$a\int_0^\infty se^{-\beta s}\mathrm{d}F(s)$$

最后由式 (6.33), 我们推出

$$g(t) = e^{-\beta t}m(t)\to(a-1)/a^2\beta\int_0^\infty se^{-\beta s}\mathrm{d}F(s)\quad(t\to\infty) \tag{6.34}$$

问题 当寿命服从参数为 λ 的指数分布 $Ex(\lambda)$ 的时候，试求 β 的值及 $g(t)$ 的极限值.

初始条件的影响 在例 6-4 中，我们假定设备在开始观测时刻 $t=0$ 之前就已开始运转. 这时一般来讲，从开始观测到第一次再生发生为止所经历时间的分布将不同于此后再生点之间所经历时间的分布. 这种随机过程称做**一般化再生过程** (generalized renewal process). 但是，若考察 $t\to\infty$ 时这一随机过程的极限性质，很显然，它不会受到初始条件的影响. 即对于一般化再生过程，式 (6.19)- 式 (6.23) 同样成立.

习 题 6

1. 试求现象发生的时间间隔服从伽玛分布 G(2,3)，即 $f(t)=4t^2e^{-2t}$ 的再生过程的均值函数 $m(t)$.

2. 如果再生过程 $\{N(t);t\geq 0\}$ 的均值函数 $m(t)=E[N(t)]=\lambda t$, 试证这个再生过程服从泊松过程.

3. 某一事务所电话打来的次数服从平均 1 分种 λ 次的泊松过程. 通话时间 $Y_1,Y_2,\cdots$ 相互独立地服从同一分布，且假定通话时电话打不进来. 现在用 $N(t)$ 表示到时刻 t 为止电话打来的次数，试证

$$\lim_{t\to\infty}\frac{E[N(t)]}{t}=\frac{\lambda}{1+\lambda E(Y_1)}$$

4. 假定有很多条线路的电话交换台电话呼叫的次数服从生起率为 λ 的泊松过程. 通话持续时间的分布函数为 G. 如果在时刻 $t=0$ 的时候没有正在通话的电话线路，而在时刻 t 的时候我们用 $m(t)$ 表示正在通话的电话线路数 $N(t)$ 的数学期望.

(1) 试证 $m(t)$ 满足再生型方程式

$$m(t) = \int_0^t [1 - G(t-s)]\lambda e^{-\lambda s}\mathrm{d}s + \int_0^t m(t-s)\lambda e^{-\lambda s}\mathrm{d}s$$

(2) 试用拉普拉斯变换解出上式，并证明

$$m(t) = \lambda \int_0^t [1 - G(s)]\mathrm{d}s$$

5. 在例 6-5 的分支过程中，假定生物体的寿命分布 F 服从连续分布，均值 μ 有限，且每一生物体留下"孩子"数的平均值 a 小于 1. 此外，还假定这种生物体绝种为止生存过的所有生物体寿命总和的数学期望为 L.

(1) 试证 $L = \int_0^\infty m(t)\mathrm{d}t$.

(2) 如果用 $\widetilde{m}(z)$ 来表示 $m(t)$ 的拉普拉斯变换，试证 $L = \widetilde{m}(0)$.

(3) 根据式 (6.29)，求出 $\widetilde{m}(z)$.

(4) 求出 L.

6. 某一设备自安装开始到发生故障为止所经历的时间服从连续分布，且其分布函数为 F. 假定这个设备或者发生故障，或者在经历时间 τ 以后马上被另一同类设备代替. 现在用 $N(t)$ 和 $M(t)$ 分别表示自安装好最初设备开始到经历时间 t 为止所替换的设备次数及所发生的故障次数.

(1) 试证 $\lim\limits_{t\to\infty} \dfrac{N(t)}{t} = 1\Big/\int_0^\tau [1 - F(u)]\mathrm{d}u$.

(2) 试证因发生故障而替换设备的时刻所组成的序列服从再生过程. 另外，试证因发生故障而替换设备的时间间隔服从如下的分布函数 G:

$$1 - G(t) = [1 - F(\tau)]^k[1 - F(t - k\tau)], \qquad (k\tau \leq t < (k+1)\tau)$$

(3) 试证 $\lim\limits_{t\to\infty} \dfrac{M(t)}{t} = F(\tau)\Big/\int_0^\tau [1 - F(u)]\mathrm{d}u$.

(4) 发生故障的时候替换设备需要 c_1 元，设备没有发生故障的时候替换设备需要 c_2 元，则单位时间平均替换费用的极限值 $C(\tau)$ 为

$$C(\tau) = \lim_{t\to\infty} \frac{c_1 M(t) + c_2[N(t) - M(t)]}{t}$$

当 F 服从指数分布的时候，试证 $C(\tau)$ 为 τ 的单调递减函数 (可见，如果在设备发生故障的时候替换设备，则所需费用变为最小).

7. (1) 试证再生过程 $\{N(t); t \geq 0\}$ 的 2 次矩 $m_2(t) = E[\{N(t)\}^2]$ 服从如下的积分方程式：

$$m_2(t) = m(t) + 2\int_0^t m(t-s)\mathrm{d}m(s)$$

(2) 试证 $m_2(t)$ 的拉普拉斯变换 $\widetilde{m}_2\ (z)$ 在 $z=0$ 的附近可写成如下形式：

$$\widetilde{m}_2\ (z) = \frac{2}{\mu^2 z^3} + \frac{2\sigma^2 - \mu^2}{\mu^3 z^2} + o(\frac{1}{z})$$

另外，当 $t \to \infty$ 的时候，试证

$$m_2(t) = \frac{t^2}{\mu^2} + \frac{2\sigma^2 - \mu^2}{\mu^3} + o(1)$$

(3) 对于任意 x, 试证下式成立：

$$\lim_{t\to\infty} P\{\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\mathrm{d}y$$

8. 考察现象发生的时间间隔分布 F 为离散型的情况，并假定 f 为 F 的概率分布. 由本章开头部分的脚注 $F(0)=0$, 令 $f(0)=0$.

(1) 试证再生方程式 (6.9) 可写成如下形式：

$$m(k) = F(k) + \sum_{j=1}^{k-1} m(k-j)f(j) \tag{6.35}$$

(2) 如果定义母函数

$$m^*(z)=\sum_{k=1}^{\infty}m(k)z^k, F^*(z)=\sum_{k=1}^{\infty}F(k)z^k, f^*(z)=\sum_{k=1}^{\infty}f(k)z^k$$

试证再生方程式可写成如下形式：

$$m^*(z)=F^*(z)+m^*(z)f^*(z) \tag{6.36}$$

(3) 试证 $F^*(z)$ 与 $f^*(z)$ 之间具有如下关系：

$$F^*(z)=\frac{f^*(z)}{1-z} \tag{6.37}$$

(4) 对于式 (6.13)，我们可以导出下式：

$$m^*(z)=\frac{f^*(z)}{1-f^*(z)}\cdot\frac{1}{1-z} \tag{6.38}$$

(5) 当现象发生的时间间隔服从几何分布 $f(k)=pq^{k-1}$. $(p+q=1)$ 的时候，试求 $m^*(z)$ 与 $m(k)$.

第7章 马尔可夫链

7.1 基本概念

7.1.1 转移概率

本章我们讨论离散时间的随机过程 $\{X_n; n = 0, 1, 2, \cdots\}$, 并假定其状态空间 S 也是离散的. 不失一般性, 令 $S = (1, 2, 3, \cdots)$. 这里, S 包含的元素个数, 即 X_n 可取的状态个数可以是有限的, 也可以是无限的.

定义 如果对于所有 $n \geq 0$ 及所有 $j \in S$,

$$P\{X_{n+1} = j | X_0, X_1, \cdots, X_n\} = P\{X_{n+1} = j | X_n\} \tag{7.1}$$

则随机过程 $\{X_n; n = 0, 1, 2, \cdots\}$ 称做马尔可夫链.[1]

式 (7.1) 从概率的角度表明了这个随机过程的"下一状态"X_{n+1} 仅依赖于"现在状态"X_n, 而与"过去状态"$X_0, X_1, \cdots, X_{n-1}$ 无关. 这种性质, 称做随机过程的**马尔可夫性**. 一般来讲, 式 (7.1) 的条件概率随着 n 的变化而变化, 但在本书我们仅讨论它与 n 无关的随机过程, 即**齐次随机过程**(homogeneous stochastic process). 这里, 令

$$P\{X_{n+1} = j | X_n = i\} = P\{i, j\} \tag{7.2}$$

[1] 准确地讲, 它应该是"离散时间"的马尔可夫链, 但我们通常简称它为马尔可夫链.

$P(i,j)$ 称做状态 i 至状态 j 的 **转移概率** (transition probability). 此外, 所有 $P(i,j)$ 进行排列得到的矩阵 $\boldsymbol{P}$,

$$\boldsymbol{P} = \begin{bmatrix} P(1,1) & P(1,2) & P(1,3) & \cdots \\ P(2,1) & P(2,2) & P(2,3) & \cdots \\ P(3,1) & P(3,2) & P(3,3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

称做 **转移概率矩阵** (transition probability matrix).

很显然, 转移概率矩阵的所有元素为非负, 且每一行的所有元素之和为 1. [1]

例 7-1(选商标) 某国销售 A,B,C,D 四种啤酒. 据某项调查表明, 消费者购买哪一种啤酒, 仅与前一次购买的啤酒种类有关, 而与这之前购买的啤酒种类无关. 现在, 用 X_0 表示某消费者最初所买啤酒的商标, 用 $X_1, X_2, \cdots$ 分别表示这之后所买啤酒的商标, 则 $\{X_n; n = 0, 1, 2, \cdots\}$ 为马尔可夫链. 其状态空间 $S = \{\text{A,B,C,D}\}$, 转移概率矩阵

$$\begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} \begin{array}{c} \begin{array}{cccc} \text{A} & \text{B} & \text{C} & \text{D} \end{array} \\ \begin{bmatrix} 0.90 & 0.05 & 0.03 & 0.02 \\ 0.10 & 0.80 & 0.05 & 0.05 \\ 0.08 & 0.10 & 0.80 & 0.02 \\ 0.10 & 0.10 & 0.10 & 0.70 \end{bmatrix} \end{array} = \boldsymbol{P}$$

在这个问题中, 我们感兴趣的是这四种啤酒的市场占有率随着时间的推移而发生的变化情况. 关于它, 在后面部分将进行具体讨论.

例 7-2(围棋赛) A,B,C 三人之间争夺围棋赛冠军, 规定最先连胜两次的人成为冠军. 最初由 A 与 B 对局, 结果 A 获胜. 接

[1] 一般来讲, 满足这两个性质的矩阵称做 **概率矩阵** (stochastic matrix).

替由 A 与 C 对局，如果 A 获胜，则 A 为冠军. 否则，再由 C 与 B 对局. 这样反复进行下去，直至决出冠军为止. 每次对局，A 赢 B, B 赢 C, C 赢 A 的概率与过去的对局结果无关，且分别为 p, q, r (假定没有平局).

上述围棋赛的过程可看做马尔可夫链，我们考虑一下它的状态空间. 如果用 AC 表示 A 在取得一次胜利之后与 C 对局的状态，那么在还没有决出冠军的时候可能具有的状态共有 AC,CB,BA 三种. 另外，我们还用 A^2,B^2,C^2 分别表示 A,B,C 为冠军时的状态. 很显然，这个马尔可夫链具有上述六种状态且其转移概率矩阵为

$$\begin{array}{c} \\ AC \\ CB \\ BA \\ A^2 \\ B^2 \\ C^2 \end{array} \begin{array}{c} \begin{array}{cccccc} AC & CB & BA & A^2 & B^2 & C^2 \end{array} \\ \left[\begin{array}{cccccc} 0 & r & 0 & 1-r & 0 & 0 \\ 0 & 0 & q & 0 & 0 & 1-q \\ p & 0 & 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array} = \boldsymbol{P} \tag{7.4}$$

决出冠军以后，这个过程也就宣结束了. 这里，需要定义所有 $\boldsymbol{P}$ 的 6×6 个元素，比如由于到达 A^2 状态以后，这一状态将持续下去，因而令 A^2 至 A^2 的转移概率为 1.

例 7-3(随机游动， random walk) 考察在直线的等间隔分布点上随机游动的质点 (见图 7-1). 质点所处的位置被称为“状态”，它可用整数表示. 如果某个时刻质点位于 i, 则下一步质点将以概率 p 向右移动一点到 $i+1$, 以概率 q 向左移动一点到 $i-1$, 或以概率 $r=1-p-q$ 停滞在原来的“状态” i. 这个质点的运动称做

(一维) 随机游动[1]，常被用于做物理学的模型或赌搏的模型．质点在直线上走动，既可以朝两方向无限移动，也可以被某一方向或两方向设置的“壁”挡住而不能继续前进．质点到达“壁”之后如果不再离开，这类壁被称为 **吸收壁** (absorbing barrier)，如果反弹回来，这类壁被称为 **反射壁** (reflecting barrier).[2] 比如，所有状态的集合 $S=\{0,1,2,3,4\}$，0 为吸收壁，4 为反射壁，那么转移概率矩阵将变为

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ \mathrm{q} & \mathrm{r} & \mathrm{p} & 0 & 0 \\ 0 & \mathrm{q} & \mathrm{r} & \mathrm{p} & 0 \\ 0 & 0 & \mathrm{q} & \mathrm{r} & \mathrm{p} \\ 0 & 0 & 0 & q' & r' \end{array}\right] \end{array} = \boldsymbol{P} \tag{7.5}$$

这里，$q'=1-r'>0$ (如果 $q'=1$，则为完全反射壁，如果 $0<q'<1$，则为不完全反射壁).

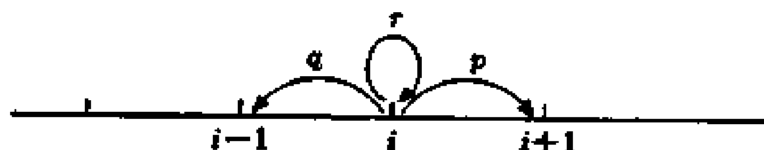


图 7-1 随机游动

例 7-4 (换灯泡) 某工厂安有 1000 个灯泡，损坏的灯泡将在月末集中替换．根据过去的统计，灯泡自安装以后的寿命 (月龄) 如表 7-1 所示．即灯泡在安装以后第一个月内的损坏率为 3%，第

[1] 若 p,q,r 随 i 发生变化，则变成更加复杂的随机游动．

[2] 移动一步，反弹回来的概率为 1，则这类反射壁为完全反射壁，否则为不完全反射壁．

二个月内的损坏率为 20%，最后到了第五个月末，所有灯泡都会损坏.

表 7-1 灯泡的替换率

经过月数	1	2	3	4	5
替换率	0.03	0.20	0.60	0.15	0.02

考察某一特定场所的灯泡，记录每月末灯泡被替换后的月龄(未满一个月，则按一个月计算)，那么它是一个马尔可夫链. 其状态空间 $S=\{1,2,3,4,5\}$，转移概率矩阵

$$P=\begin{bmatrix} p_1 & q_1 & 0 & 0 & 0 \\ p_2 & 0 & q_2 & 0 & 0 \\ p_3 & 0 & 0 & q_3 & 0 \\ p_4 & 0 & 0 & 0 & q_4 \\ p_5 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} & p_i+q_i=1(1\leq i\leq 4) \\ & p_5=1 \end{aligned} \tag{7.6}$$

于是，根据 p_i, q_i 与表 7-1 灯泡替换率之间存在的以下关系. 我们能够算出 p_i, q_i.

$$\begin{aligned} p_1 &= 0.03 \\ q_1p_2 &= 0.20 \\ q_1q_2p_3 &= 0.60 \\ q_1q_2q_3p_4 &= 0.15 \\ q_1q_2q_3q_4p_5 &= 0.02 \end{aligned}$$

7.1.2 切普曼 - 科尔莫哥洛夫方程式(Chapman-Kolmogorov)

我们知道 $P(i,j)$ 是某一时刻 n 位于状态 i 的马尔可夫链在下一时刻移动到状态 j 的概率. 现在假定 $X_n=i$，然后考虑“下下个

时刻" $n+2$ 时移动到状态 j 的概率

$$P\{X_{n+2}=j|X_n=i\}$$

为了算出这个概率，我们需要考察它在 $n+1$ 时刻所处的状态 (如图 7-2). 那么，根据全概率公式.

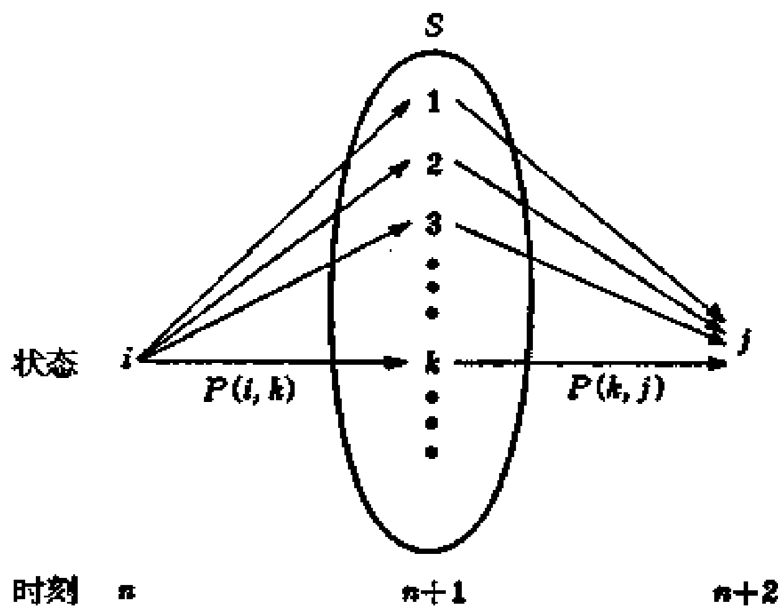


图 7-2

$$\begin{aligned}P\{X_{n+2}=j|X_n=i\}&=\sum_{k\in S}P\{X_{n+1}=k|X_n=i\}P\{X_{n+2}\\&=j|X_{n+1}=k\}\\&=\sum_{k\in S}P(i,k)P(k,j)\end{aligned} \tag{7.7}$$

很显然，式 (7.7) 的右边相当于 P 乘以 P 后得到的矩阵的 (i,j) 元素. 同理，我们可以计算三步及三步以上的转移概率，即如果令

$$P^{(m)}(i,j)=P\{X_{n+m}=j|X_n=i\} \tag{7.8}$$

为矩阵 $\boldsymbol{P}^{(m)}$ (m 步转移概率矩阵, m-step transition probability matrix) 的 (i,j) 元素，则对于任意 $m \geq 1$,

$$\boldsymbol{P}^{(m)} = \boldsymbol{P}^m \tag{7.9}$$

特别地， $i = j$ 的时候定义 $P^{(0)}(i,j) = 1$, $i \neq j$ 的时候定义 $P^{(0)}(i,j) = 0$, 那么 $m = 0$ 的时候式 (7.9) 仍然成立.

对于任意 $l(0 \leq j \leq m)$, 由矩阵的性质

$$\boldsymbol{P}^m = \boldsymbol{P}^l \boldsymbol{P}^{m-l}$$

故由式 (7.9)

$$\boldsymbol{P}^{(m)} = \boldsymbol{P}^{(l)} \boldsymbol{P}^{(m-l)} \tag{7.10}$$

即

$$P^{(m)}(i,j) = \sum_{k \in S} P^{(l)}(i,k) P^{(m-l)}(k,j) \quad (i,j \in S) \tag{7.11}$$

上式被称为**切普曼 - 科尔莫哥若夫方程式**.

问题 对于任意 $m \geq 0$, 试证 $\boldsymbol{P}^{(m)}$ 为概率矩阵.

7.2 状态的分类及其性质

7.2.1 等价类

如果对于状态 i 和 j, 存在某个 $n \geq 0$, 使得 $P^{(n)}(i,j) > 0$, 则称这个马尔可夫链**自状态 i 可达状态 j**(accessible), 并记为 $i \to j$. 从图论的角度,我们还可进行如下解释. 如图 7-3, 某**转移图** (transition graph) 的每个结点对应于马尔可夫链的每个状态，从某个结点 i 指向另一个结点 j 的有向边对应于满足 $P(i,j) > 0$ 的所有状态组合 (i,j). 如果图中存在一条或一条以上从结点 i 至结点 j 的**有向**

边 (oriented path), 我们称自状态 i 可达状态 j. 这里, 假定自某一状态总可达到自身状态.

如果自状态 i 可达状态 $j(i \to j)$, 且自状态 j 可达状态 $i\ (j \to i)$, 则称 **状态 i 和状态 j 相通** (communicate), 并记为 $i \leftrightarrow j$. 在有向图中, 它意味着既存在从结点 i 到结点 j 的有向边, 也存在从结点 j 到结点 i 的有向边. 我们容易推出下面定理.

定理 7-1 相通关系是等价关系 (equivalence relationship). 即它满足

(1) (自反律) $i \leftrightarrow i$

(2) (对称律) 如果 $i \leftrightarrow j$, 则 $j \leftrightarrow i$.

(3) (传递律) 如果 $i \leftrightarrow j$ 且 $j \leftrightarrow k$, 则 $i \leftrightarrow k$.

利用等价关系 $\leftrightarrow$, 我们可以把所有状态按等价类 (唯一地) 进行划分. 任意两个属于同一类的状态之间是相通的, 任意两个属于不同类的状态之间是不相通的 (后面的场合, 也可能出现单向可达的情况). 当所有状态被划分为同一类的时候, 我们称此马尔可夫链是 **既约的** (irreducible). 图论中, 我们称这种有向图是 **强连通的** (strongly connected).

例 7-5 图 7-3 给出了例 7-3 随机游动的有向图 (这里, $r_4 > 0$). $\{1, 2, 3, 4\}$ 是一个等价类, $\{0\}$ 是另一个等价类.

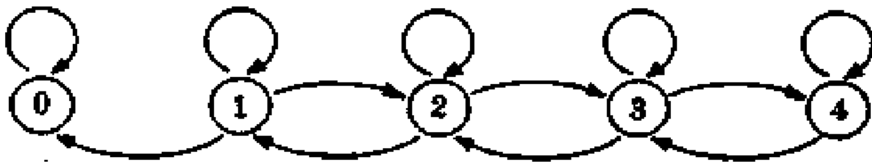


图 7-3 随机游动的转移图

问题 试给出例 7-1 及例 7-2 马尔可夫链的有向图，并将其状态按等价类划分.

属于同一等价类的所有状态有许多共通的性质. 下面讲一讲这些性质.

7.2.2 周期

满足 $P^{(n)}(i,i)>0$ 的所有 $n\geq 1$ 的整数的最大公约数被称为这个马尔可夫链的关于状态 i 的周期 $d(i)$ (如果不存在这样的 n, 则定义 $d(i)=0$), 周期为 1 的状态被称为是 **非周期的** (aperiodic).

定理 7-2 如果 $i\leftrightarrow j$, 则 $d(i)=d(j)$.

这个定理的证明，参照习题 7 的问题 5.

7.2.3 常返性

马尔可夫链的状态，既可能是依概率 1 无数次返回的，也可能不是这样的. 比如在图 7-3 有向图所示的马尔可夫链中，状态 0 属于前者，状态 1,2,3,4 属于后者. 下面，我们详细考察这类性质.

系统首次到达状态 j 的时刻 $n(\geq 1)$, 称为 **首次进入状态 j 的时刻** (first passage time), 并记为 T_j. 即

如果 $X_1\neq j, X_2\neq j,\cdots,X_{n-1}\neq j, X_n=j$, 则 $T_j=n$.

如果对于所有 $n(n\geq 1)$, 存在 $X_n\neq j$, 则 $T_j=\infty$.

那么,

$$f^{(n)}(i,j)=P\{T_j=n|X_0=i\}\quad (n\geq 1)\tag{7.12}$$

表示系统从状态 i 出发首次进入状态 j 的时刻为 n 的概率.

$$f(i,j) = \sum_{n=1}^{\infty} f^{(n)}(i,j) \tag{7.13}$$

表示系统从状态 i 出发进入状态 j 的概率.[1]

如果 $f(j,j) = 1$, 则称状态 j 是 **常返的** (recurrent), 如果 $f(j,j) < 1$, 则称状态 j 是 **非常返的** (nonrecurrent). 如用 T_j 表示, 则 $P\{T_j = \infty | X_0 = j\} = 0$ 的时候是常返的; $P\{T_j = \infty | X_0 = j\} > 0$ 的时候是非常返的.

$f^{(n)}(i,j)$ 与 $P^{(n)}(i,j)$ 之间有以下关系.

$$P^{(n)}(i,j) = \sum_{k=1}^{n} f^{(k)}(i,j) P^{(n-k)}(j,j) \tag{7.14}$$

下面验证上式.

$P^{(n)}(i,j)$ 是 $X_0 = i$ 条件下 $X_n = j$ 的概率. 而事件 $\{X_n = j | X_0 = i\}$ 可以表示成为 T_j 的互不相容事件之和:

$$\{X_n = j | X_0 = i\} = \bigcup_{k=1}^{n} \{X_n = j, \quad T_j = k | X_0 = i\}$$

故由全概率公式 (1.21),

$$\begin{aligned} P\{X_n = j | X_0 = i\} &= \sum_{k=1}^{n} P\{T_j = k | X_0 = i\} P\{X_n = j | T_j = k, X_0 = i\} \\ &= \sum_{k=1}^{n} P\{T_j = k | X_0 = i\} P\{X_n = j | X_k = j\} \\ &= \sum_{k=1}^{n} f^{(k)}(i,j) P^{(n-k)}(j,j) \end{aligned}$$

---

[1] 很显然, $\sum_{n=1}^{\infty}$ 相当于 $\lim_{m \to \infty} \sum_{n=1}^{m}$. 下同.

这里的证明方法，首先是把事件划分为 T_j 的互不相容事件之和，然后通过全概率公式求出这个事件的概率. 今后我们经常用到这种证明方法，现简称它为“停时方法”.

如果状态 j 是常返的，则系统依概率 1 无数次返回状态 j; 如果状态 j 是非常返的，则系统依概率 0 无数次返回状态 j. 直观地，这很明显，但从理论上我们需要如下证明.

用随机变量 N_j 表示 $n \geq 1$ 的时候，系统进入状态 j 的次数，并令

$$g(i,j) = P\{N_j = \infty | X_0 = i\} \tag{7.15}$$

根据“停时方法”，

$$\begin{aligned}
P\{N_j \geq n | X_0 = i\} &= \sum_{k=1}^{\infty} P\{T_j = k | X_0 = i\} P\{N_j \geq n | T_j = k, X_0 = i\} \\
&= \sum_{k=1}^{\infty} P\{T_j = k | X_0 = i\} P\{N_j \geq n | X_k = j\} \\
&= \sum_{k=1}^{\infty} P\{T_j = k | X_0 = i\} P\{N_j \geq n-1 | X_0 = j\} \\
&= f(i,j) P\{N_j \geq n-1 | X_0 = j\}
\end{aligned} \tag{7.16}$$

反复使用上式，最后可推出

$$\begin{aligned}
P\{N_j \geq n | X_0 = i\} &= f(i,j)\{f(j.j)\}^{n-2} P\{N_j \geq 1 | X_0 = j\} \\
&= f(i,j)\{f(j,j)\}^{n-1} \quad (n \geq 1)
\end{aligned} \tag{7.17}$$

两边取极限，

$$g(i,j) = f(i,j) \lim_{n\to\infty} \{f(j,j)\}^{n-1} \tag{7.18}$$

这里，令 $i = j$,

如果 j 是常返的，则 $(f(j,j) = 1) \Leftrightarrow g(j,j) = 1$

如果 j 是非常返的，则 $(f(j,j) < 1) \Leftrightarrow g(j,j) = 0$

由于马尔可夫链的构造是由转移概率矩阵 P 决定的，所以我们可以直接通过 P 的元素判定某一状态是否常返. 关于它，有以下定理.

定理 7-3 状态 j 常返的充要条件是

$$\sum_{n=1}^{\infty} P^{(n)}(j,j) = \infty \tag{7.19}$$

证明 首先引入随机变量 Z_n, 使得在 $X_0 = i$ 的条件下如果 $X_n = j$ 则其取值为 1, 否则其取值为 0. 那么 $\sum_{n=1}^{\infty} Z_n$ 相当于从状态 i 出发的系统到达状态 j 的次数. 由于

$$E[\sum_{n=1}^{\infty} Z_n] = \sum_{n=1}^{\infty} E[Z_n] = \sum_{n=1}^{\infty} P\{Z_n = 1\} = \sum_{n=1}^{\infty} P^{(n)}(i,j) \tag{7.20}$$

我们知道式 (7.19) 的条件表示从状态 j 出发的系统返回状态 j 的次数的数学期望为无穷大. 另一方面，从式 (7.17) 及定理 3-2 的推论，我们还知道

$$\begin{aligned} E[\sum_{n=1}^{\infty} Z_n] &= \sum_{n=1}^{\infty} P\{N_j \geq k|X_0 = i\} \\ &= f(i,j) \sum_{k=1}^{\infty} \{f(j,j)\}^{k-1} \end{aligned} \tag{7.21}$$

故在式 (7.20) 及式 (7.21) 中令 $i = j$ 便可以推出

$f(j,j) = 1$ 等同于 $\sum_{n=1}^{\infty} P^{(n)}(j,j) = \infty$.

这个定理告诉我们常返性是等价类所特有的性质. [1]

推论 1 同一等价类的所有状态，或者全部是常返的，或者全部是非常返的.

证明 如果我们能够从 $i \leftrightarrow j$ 及 $\sum_{m=1}^{\infty} P^{(m)}(i,i) = \infty$ 推出 $\sum_{m=1}^{\infty} P^{(m)}(j,j) = \infty$ 即可.

如果 $i \leftrightarrow j$, 则存在某个 $l \geq 1$, $n \geq 1$ 使得 $P^{(l)}(j,i) > 0$, $P^{(n)}(i,j) > 0$. 由转移概率的概念容易推出不等式

$$P^{(l+m+n)}(j,j) \geq P^{(l)}(j,i)P^{(m)}(i,i)P^{(n)}(i,j) \quad (m = 1, 2, \cdots)$$

即

$$\sum_{m=1}^{\infty} P^{(l+m+n)}(j,j) \geq P^{(l)}(j,i)\{\sum_{m=1}^{\infty} P^{(m)}\}(i,i)\}P^{(n)}(i,j)$$

于是，如果 i 是常返的，则其右边为 ∞，左边也为 ∞, 即 j 是常返的.

推论 2 如果状态 j 是非常返的，则对于任意状态 $j \in S$,

$$\lim_{n \to \infty} P^{(n)}(i,j) = 0 \tag{7.22}$$

证明 因为 $0 \leq f(j,j) < 1$, 故由式 (7.20) 及式 (7.21)

$$\sum_{n=1}^{\infty} P^{(n)}(i,j) = \frac{f(i,j)}{1 - f(j,j)} < \infty \tag{7.23}$$

[1] 因此，今后我们可以说“等价类是常返的 (或非常返的)”.

即式 (7.22) 成立.

例 7-6(随机游动；继续) 在例 7-3 随机游动中，我们考虑质点可向左右两方向无限移动且 $r=0$, 即质点在一次移动中必然要走到左右相邻的某一状态的情况. 很显然，如果 $p>0, q>0$, 则这个系统是既约的. 现在，考察状态 0 的常返性.

很明显，

$$P^{(2n+1)}(0,0)=0 \qquad (n=0,1,2,3,\cdots)$$
$$P^{(2n)}(0,0)=\begin{pmatrix} 2n \\ n \end{pmatrix} p^n q^n=\frac{(2n)!}{(n)!^2} p^n q^n \quad (n=1,2,3,\cdots)$$

由斯特林公式. (Stirling formula)

$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$$
$$p^{(2n)}(0,0) \sim \frac{(4pq)^n}{\sqrt{\pi n}}$$

这里，等号仅在 $4pq \leq 1$, $p=q=1/2$ 的情况下成立.

可见，当且仅当 $p=q=1/2$ 的情况下，

$$\sum_{n=1}^{\infty} P^{(n)}(0,0)=\infty$$

即这个系统是常返的.

同理，二维空间格子点上对称的随机游动 (在每一次移动中走到相邻四个格子点的概率均为 1/4) 所对应的系统也是常返的. 但对于三维空间来讲，对称的随机游动则是非常返的，即质点离开某一点后不再返回这一点的概率是正的.

定理 7-3 描述了同一等价类状态之间的关系，下面讲一讲不同等价类状态之间的关系.

定义 用 C 来表示马尔可夫链的任意非空状态集合. 当从任

意属于 C 的状态都不能到达某一不属于 C 的状态的时候，我们称集合 C 是 **闭的** (closed).[1]

定理 7-4 常返的等价类是闭的.

证明 设 i 是常返的状态，j 是任意从 i 可达的状态. 则从状态 i 出发不返回 i 而到达 j 的概率为 p. 又在到达 j 的条件下不返回 i 的概率为 $1-f(j,i)$. 故从状态 i 出发不返回 i 的概率满足不等式

$$1-f(i,i) \ge p\{1-f(j,i)\} \ge 0$$

由于 i 是常返的，所以 $1-f(i,i)=0$, 又由于 $p>0$, 所以 $f(j,i)=1$, 最后我们得到 $j \to i$. 即 j 与 i 属于同一等价类.

推论 如果 i 和 j 属于同一常返的等价类，则

$$f(i,j)=g(i,j)=1 \tag{7.24}$$

证明 在式 (7.16) 的两边取极限，则

$$g(i,j)=f(i,j)g(j,j) \tag{7.25}$$

再由定理 7-4 的证明便可推出结论.

平均到达时间 (mean first passage time), **平均返回时间** (mean recurrence time) 我们用 $m(i,j)$ 表示从状态 i 出发首次到达状态 j 的系统所经历时间的数学期望，即如果 $P\{T_j<\infty|X_0=i\}=1$,

---

[1] 注意这里所讲的闭的概念与一般生活中闭的概念的区别，即从 C 的外部进入 C 是可能的，比如我们可以想象抓耗子的捕鼠器.

则

$$m(i,j) = \sum_{n=1}^{\infty} nP\{T_j = n|X_0 = i\}$$
$$= \sum_{n=1}^{\infty} nf^{(n)}(i,j) \tag{7.26}$$

否则

$$m(i,j) = \infty \tag{7.27}$$

一般地，$m(i,j)$ 称做状态 i 到状态 j 的 **平均到达时间**. 特别地，$m(j,j)$ 称做状态 j 的 **平均返回时间**.

如果状态 j 是非常返的，则因为 $P\{T_j = \infty|X_0 = j\} > 0$, 故 $m(j,j) = \infty$. 如果状态 j 是常返的，则 $P\{T_j = \infty|X_0 = j\} = 0$, 但式 (7.26) 的取值既可能是有限的，也可能是无限的. 如果 $m(j,j) < \infty$, 则状态 j 是 **正常返的** (positive recurrent), 如果 $m(j,j) = \infty$, 则状态 j 是 **零常返的** (null recurrent).[1] 这里，我们给出以下定理.

定理 7-5 (1) 属于同一常返等价类的状态或者都是正常返的，或者都是零常返的.

(2) 如果属于某一常返等价类的状态个数是有限的，则这个等价类是正常返的.

平均到达时间，平均返回时间的计算方法 根据定义式 (7.27), 我们可能不易算出 $m(i,j)(i \neq j)$ 或 $m(j,j)$ 的值. 这时候利用下面的式 (7.28) 会方便得多.

考虑从状态 i 出发到达状态 j 的过程中，第一步将转移到哪

[1] "正" 或 "零" 对应于 $1/m(j,j)$ 的正或零. 而 $1/m(j,j)$, 正如后面所讲，是指系统状态在很长一段时间内处于状态 j 的比率的数学期望.

一状态，即

$$
\begin{aligned}
m(i,j) &= P(i,j)\times 1+\sum_{k\neq j}P(i,k)\{1+m(k,j)\} \\
&= 1+\sum_{k\neq j}P(i,k)m(k,j) \qquad (7.28)
\end{aligned}
$$

固定 j, 式 (7.28) 将变为 $m(i,j)(j\in S-\{j\})$ 的联立方程式. 此外，如果令 $i=j$, 则式 (7.28) 将变为平均常返时间的计算公式.

例 7-7(随机游动) 下面考虑图 7-4 所示转移概率在非负整数全体 $\{S=0,1,2,\cdots\}$ 上的随机游动. 这里，$0<p=1-q<1$. 从图中可以看出所有状态是相通的，即属于同一等价类. 此外，如果 $q\geq p$, 则是常返的，如果 $q<p$, 则是非常返的. 现在我们考虑在常返的情况下，它是正常返的还是零常返.



图 7-4 非负整数全体随机游动的转移概率

在式 (7.28) 中，令 $j=0$, $i=1,2,\cdots$, 则我们可以得到有关状态 i 到状态 $j=0$ 平均到达时间 $m(i,0)$ 的联立方程式.

$$
m(1,0)=1+pm(2,0) \qquad (7.29a)
$$

$$
m(i,0)=1+pm(i+1,0)+qm(i-1,0) \quad (i\geq 2) \qquad (7.29b)
$$

式 (7.29b) 是 $m(i,0)(i\geq 1)$ 的二阶差分方程，我们可以借助

于差分方程的一般解法去解它[1]. 式 (7.29b) 的特征方程式为

$$\lambda = p\lambda^2 + q$$

又 $p+q=1$, 所以其特征根为 $\lambda_1 = 1$ 及 $\lambda_2 = q/p$. 故 $\lambda_1 \neq \lambda_2$ 的时候 (即 $p<q$), 式 (7.29b) 的解为

$$m(i,0) = A\lambda_1^i + B\lambda_2^i + C_i = A + B(q/p)^i + C_i \tag{7.30}$$

$\lambda_1 = \lambda_2$ 的时候 (即 $p=q$), 式 (7.29b) 的解为

$$m(i,0) = (A' + B'i)\lambda_1^i + C_i' = A' + B'i + C_i' \tag{7.31}$$

这里, A, B, A', B' 是不依赖于 i 的常数, C_i, C_i' 是式 (7.29b) 的特解.

特解的求法与微分方程式的情况一样, 即依次令 $m(i,0)$ 为常数, 一次式, 二次式, $\cdots$, 并把它代入到式 (7.29b) 中去, 求出其的解即可. 于是,

如果 $p<q$, 我们可以求出特解 $\quad C_i = \dfrac{i}{q-p}$

如果 $p=q$, 我们可以求出特解 $\quad C_i' = -i^2$

这里 A, B 等常数, 可如下求出.

式 (7.29a) 相当于在式 (7.29b) 中令 $i=1, m(0,0)=0$. 如果 $i=0$ 的时候, 式 (7.30) 或式 (7.31) 的右边取值为 0, 那么式 (7.29a) 的条件可看做是式 (7.29b) 的边界条件. 于是, $p<q$ 的时候, $B=-A$; $p=q$ 的时候, $A'=0$, 其解变为

$$m(i,0) = \begin{cases} A\{1-(q/p)^i\} + i/(q-p) & (p<q) \quad (7.32a) \\ B'i - i^2 & (p=q) \quad (7.32b) \end{cases}$$

[1] 完全类同于二阶线形常微分方程的解法. 不了解差分方程论的读者可参照微分方程论去思考.

下面考察一下 A 的取值. 从图中我们看出 p 趋近于 0, q 趋近于 1 的时候, $m(i,0)$ 趋近于 i. 可是在式 (7.32a) 中, 如果 $A \neq 0$, 则 $m(i,0)$ 的取值将无限远离 i. 因此, A 只能等于 0. 另一方面, 如果 B' 的值为有限, 则在式 (7.32b) 中令 i 为无穷大的时候, $m(i,0)$ 将变为负值. 所以, B' 也只能等于 ∞ . 于是,

$$m(i,0) = \begin{cases} i/(q-p) & (p < q) \qquad (7.33a) \\ \infty & (p = q) \quad (i \geq 1) \qquad (7.33b) \end{cases}$$

在式 (7.28) 中令 $i = j = 0$, 我们得到

$$m(0,0) = 1 + pm(1,0)$$

再根据上面结果, 我们得到状态 0 的平均返回时间

$$m(0,0) = \begin{cases} q/(p-q) & (p < q) \qquad (7.34a) \\ \infty & (p = q) \qquad (7.34b) \end{cases}$$

很显然, $p < q$ 的时候, 这个系统是正常返的, $p = q$ 的时候, 这个系统是零常返的.

7.3 吸收概率和平均吸收时间

7.3.1 吸收概率

对于同时存在常返状态和非常返状态的马尔可夫链, 我们感兴趣的是从特定的非常返状态 i 出发到达特定的常返状态 j 的概率. 这个概率可被看作是从状态 i 出发的系统被状态 j 所属的等价类“吸收”的概率. 它可通过下面的联立方程式求出.

$$f(i,j) = \sum_{k \in C(j)} P(i,k) + \sum_{k \in T} P(i,k) f(k,j) \quad (i \in T) \qquad (7.35)$$

这里，$C(j)$ 是状态 j 所属的常返的等价类，T 是非常返的状态集合．下面验证此式．假定从状态 i 出发的系统转移一步以后到达状态 k, 则

$$
\begin{aligned}
f(i,j) &= \sum_{k\in S} P(i,k)f(k,j) \\
&= \sum_{k\in C(j)} P(i,k)f(k,j) + \sum_{k\in T} P(i,k)f(k,j) \\
&\quad + \sum_{k\in S-C(j)-T} P(i,k)f(k,j)
\end{aligned}
$$

如果 $k \in C(j)$, 则 $f(k,j) = 1$(定理 [7-4] 的推论);

如果 $k \in S - C(j) - T$, 则 $f(k,j) = 0$ (定理 [7-4]).

由此我们推出式 (7.35).

例 7-8 在例 7-2 的围棋赛中，我们感兴趣的是每个人获得冠军的概率．下面求一求 A 获冠军的概率．这时，式 (7.35) 中的 j 相当于状态 A^2, $C(j)$ 也相当于状态 A^2. 此外，$T = \{\mathrm{AC},\mathrm{CB},\mathrm{BA}\}$. 因而式 (7.35) 相当于联立方程式

$$
\begin{cases}
f(\mathrm{AC},\mathrm{A}^2) = & (1-r) + rf(\mathrm{CB},\mathrm{A}^2) \\
f(\mathrm{CB},\mathrm{A}^2) = & qf(\mathrm{BA},\mathrm{A}^2) \\
f(\mathrm{BA},\mathrm{A}^2) = & pf(\mathrm{AC},\mathrm{A}^2)
\end{cases}
$$

解之，即得到了 A 获冠军的概率

$$
f(\mathrm{AC},\mathrm{A}^2) = \frac{1-r}{1-pqr}
$$

问题 试求 B, C 获冠军的概率．

7.3.2 有限马尔可夫链 (finite Markov chain)的平均吸收时间

下面考察状态集合 S 有限的马尔可夫链．对于 S 来讲，T 表示非常返的状态集合 (假定 T 为非空集合．因为在有限马尔可

夫链中至少存在一个常返的状态，所以 $S-T$ 也为非空集合). 此外，$m(i)$ 表示从状态 $i \in T$ 出发首次到达属于 $S-T$ 的某状态时所经历时间的均值，我们把它记做 **平均吸收时间** (mean time to absorption). 仿式 (7.28), $m(i)$ 满足下式

$$m(i) = 1 + \sum_{j \in T} P(i,j)m(j) \tag{7.36}$$

如果用矩阵和向量表示，则

$$\boldsymbol{m} = \boldsymbol{1} + \boldsymbol{Q}\boldsymbol{m} \tag{7.37}$$

这里，列向量 $\boldsymbol{m}$ 的元素是 $m(i)(i \in T)$, 列向量 $\boldsymbol{1}$ 的所有元素是 1, $\boldsymbol{Q}$ 是转移矩阵 $\boldsymbol{P}$ 中除去 T 状态对应的行和列得到的矩阵. 即如果用从小到大的标号去标记属于 T 的状态和不属于 T 的状态，那么 $\boldsymbol{P}$ 可写成

$$\begin{array}{c} \\ T \\ S-T \end{array}\begin{array}{c} \begin{array}{cc} T & S-T \end{array} \\ \left[\begin{array}{c|c} \boldsymbol{Q} & \boldsymbol{R} \\ \hline \boldsymbol{0} & \boldsymbol{U} \end{array}\right] \end{array} = \boldsymbol{P} \tag{7.38}$$

的形式. 这里，矩阵 $\boldsymbol{U}$ 表示常返的状态集合 $S-T$ 之间的转移概率，矩阵 $\boldsymbol{R}$ 表示从非常返的状态转移到常返状态的概率. 很显然，我们容易推出 $\boldsymbol{P}$ 的 n 乘

$$\boldsymbol{P}^n = \left[\begin{array}{c|c} \boldsymbol{Q}^n & \boldsymbol{R}_n \\ \hline \boldsymbol{0} & \boldsymbol{U}^n \end{array}\right]$$

(根据数学归纳法). 这里，$\boldsymbol{Q}^n$ 与 $\boldsymbol{U}^n$ 分别表示 $\boldsymbol{Q}$ 与 $\boldsymbol{U}$ 的 n 乘，而 $\boldsymbol{R}_n$ 则不表示 $\boldsymbol{R}$ 的 n 乘而是一个与 $\boldsymbol{Q}$ 及 $\boldsymbol{U}$ 有关的量. 如果 i 和 j 是非常返的状态，则正如定理 7-3 推论 2 所证，

$$\sum_{n=1}^{\infty} \boldsymbol{Q}^n(i,j) = \sum_{n=1}^{\infty} f(i,j)\{f(j,j)\}^{n-1}$$

将会收敛．因而，如果假定 I 和 Q 为大小相同的单位矩阵，则存在 $I-Q$ 的逆矩阵 N，使得

$$N=(I-Q)^{-1}=I+Q+Q^2+Q^3+\cdots \tag{7.39}$$

N 有时也被称为有限马尔可夫链的**基础矩阵**(fundamental matrix)．最后，我们得到式 (7.37) 的解

$$m=N\mathbf{1} \tag{7.40}$$

例 7-9 在例 7-2 中，我们求一求直至决出冠军为止所进行比赛次数的数学期望．这里，$T=\{\mathrm{AC},\mathrm{CB},\mathrm{BA}\}\equiv\{1,2,3\}$，

$$Q=\begin{bmatrix}0 & r & 0\\ 0 & 0 & q\\ p & 0 & 0\end{bmatrix}$$

于是，对应于式 (7.37)，

$$\begin{cases}m(1)=1+rm(2)\\ m(2)=1+qm(3)\\ m(3)=1+p(m(1)\end{cases}$$

解之，

$$m(1)=\frac{1+r+rq}{1-pqr}$$

这一题，通过 N 计算 $m(1)$ 并非上策，但正如后面所述，一般来讲 N 也还要用在其它方面，故不妨用它来算算看．根据矩阵的性质，

$$Q^2=\begin{bmatrix}0 & 0 & rq\\ qp & 0 & 0\\ 0 & pr & 0\end{bmatrix}\qquad Q^3=\begin{bmatrix}pqr & 0 & 0\\ 0 & pqr & 0\\ 0 & 0 & pqr\end{bmatrix}$$

于是

$$N=\sum_{n=0}^{\infty}Q^n=\sum_{n=0}^{\infty}Q^{3n}(I+Q+Q^2)=\sum_{n=0}^{\infty}(pqr)^n(I+Q+Q^2)$$
$$=\frac{1}{1-pqr}(I+Q+Q^2)$$

这里，

$$I+Q+Q^2=\begin{bmatrix}1 & r & rq\\ qp & 1 & q\\ p & pr & 1\end{bmatrix}$$

7.4 转移概率的极限定理

正如定理 7-3 的推论 2 所述，如果状态 j 是非常返的，则对于任意 $i\in S$, $n\to\infty$ 的时候，$P^{(n)}(i,j)\to 0$. 但如果状态 j 是常返的，$P^{(n)}(i,j)$ 却不一定收敛于某一极限. 下面的定理，叙述了 $n\to\infty$ 时 $P^{(n)}(i,j)$ 的变化情况及它与状态 j 的平均返回时间的关系. 这部分内容在下一节的讨论中将起重要作用.

定理 7-6 如果状态 j 是正常返的，则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n}P^{(l)}(j,j)=\frac{1}{m(j,j)} \tag{7.41}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n}P^{(l)}(i,j)=f(i,j)\times\frac{1}{m(j,j)}\quad(\text{对于任意 } i\in S) \tag{7.42}$$

如果状态 j 是零常返的，则对于任意 i,

$$\lim_{n\to\infty}P^{(n)}(i,j)=0 \tag{7.43}$$

这里，我们省略定理的详细证明，下面简单地通过再生过程理论理解这个定理. 把离散时间的系统看作连续时间的再生过程，把系统在时刻 n 所处的状态 $k(X_n = k)$ 看作在时间 $n \le t < n+1$ 再生过程处于状态 k. 另外，把系统进入状态 j 的时刻看作再生过程中的再生点，并在这一再生过程 $\{N(t); t \ge 0\}$ 中，令 $E[N(t)] = m(t)$.

正如 [定理 7-3] 证明所述，$\sum_{l=1}^{n} P^{(l)}(j,j)$ 表示从状态 j 出发的系统，在 n 步以内返回状态 j 所需转移次数的数学期望. 于是

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P^{(l)}(j,j) = \lim_{t\to\infty} \frac{m(t)}{t}$$

最后根据再生过程的极限公式 (6.19) 得到式 (7.41).

下面再考察式 (7.42). 仿上，$R_n \equiv \sum_{l=1}^{n} P^{(l)}(i,j)\sqrt{n}$ 表示从 $X_0 = i$ 出发的系统，在 n 步以内返回状态 j 所需转移次数的数学期望. 则系统首次进入状态 j 所需时间 T_j 有限的时候 (其概率为 $f(i,j)$)，根据第六章有关 (一般再生过程的内容)，$\lim_{n\to\infty} R_n$ 相当于 $\lim_{t\to\infty} m(t)/t$. 又由于 $T_j = \infty$ 的时候，$\lim_{n\to\infty} R_n = 0$. 所以式 (7.42) 成立. 最后在式 (6.22) 中令 $\mu = m(j,j) = \infty$，我们可能得到式 (7.43).

7.5 平稳分布和极限分布

这一节我们将讨论经过很长一段时间以后，系统状态概率分布的变化趋势. 首先，重新回顾一下换灯泡问题 (例 7-4).

例 7-10 (换灯泡；继续) 这个马尔可夫链的转移概率矩阵.

$$
\boldsymbol{P}=\begin{bmatrix} p_1 & q_1 & 0 & 0 & 0 \\ p_2 & 0 & q_2 & 0 & 0 \\ p_3 & 0 & 0 & q_3 & 0 \\ p_4 & 0 & 0 & 0 & q_4 \\ p_5 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{7.44}
$$

我们用 $v_n(i)$ 表示自观测开始到第 n 个月替换之后处于状态 i 的灯泡个数的数学期望，那么

$$
\begin{aligned}
v_n(1) &= p_1v_{n-1}(1) + p_2v_{n-1}(2) + p_3v_{n-1}(3) + p_4v_{n-1}(4) + p_5v_{n-1}(5) \\
v_n(2) &= q_1v_{n-1}(1) \\
v_n(3) &= q_2v_{n-1}(2) \\
v_n(4) &= q_3v_{n-1}(3) \\
v_n(5) &= q_4v_{n-1}(4)
\end{aligned} \tag{7.45}
$$

再用行向量 $\boldsymbol{v}_n$ 表示上式左边出现的 5 个数学期望. 则式 (7.65) 可简写成

$$
\boldsymbol{v}_n = \boldsymbol{v}_{n-1}\boldsymbol{P} \qquad (n \geq 1) \tag{7.46}
$$

如果知道初始条件 $\boldsymbol{v}_0$, 即知道观测开始时候不同月龄的灯泡个数，则可通过式 (7.45) 或 (7.46) 依次求出. $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots$.

表 7-2 给出了两种初始条件下的一部分计算结果. 从这个表可以看出，初始条件的影响将逐渐变弱，且它慢慢逼近同一月龄分布. 一般地，下面的讨论告诉我们无论初始条件怎样，它都会逼近同一月龄分布. 又由于灯泡总数恒为 1000 个，所以 $v_n(i)/1000$ 相当于最初 ($n=0$ 时) 选定的任一场所的灯泡在第 n 个月的月末

表 7-2 灯泡月龄分布 (数学期望) 的变化 (其 1).

n	$\nu_n(1)$	$\nu_n(2)$	$\nu_n(3)$	$\nu_n(4)$	$\nu_n(5)$
0	1000	0	0	0	0
1	30	970	0	0	0
2	201	29	770	0	0
3	612	195	23	170	0
4	226	594	155	5	20
5	274	220	471	34	1
6	452	266	174	104	4
7	300	438	211	39	12
8	310	291	348	47	4
9	386	301	231	77	5
10	327	374	239	51	9
11	327	317	297	53	6
12	359	317	252	66	6
13	336	348	252	56	8
14	335	326	277	56	6
15	348	325	259	61	7
16	340	338	258	57	7
17	338	330	268	57	7
18	344	328	262	59	7
19	341	334	261	58	7
20	340	331	265	57	7
21	342	330	262	59	7
22	341	332	262	58	7
23	341	331	264	58	7
24	342	330	263	58	7

灯泡月龄分布 (数学期望) 的变化 (其 2)

n	$\nu_n(1)$	$\nu_n(2)$	$\nu_n(3)$	$\nu_n(4)$	$\nu_n(5)$
0	500	300	200	0	0
1	233	485	238	44	0
2	332	226	385	53	5
3	408	322	179	85	6
4	299	396	255	40	10
5	334	290	314	56	5
6	369	324	231	69	7
7	325	358	257	51	8
8	337	316	284	57	6
9	353	327	250	63	7
10	335	342	260	55	7
11	339	325	272	57	7
12	346	329	258	60	7
13	339	336	261	57	7
14	340	329	266	58	7
15	343	330	261	59	7
16	340	333	262	58	7
17	341	330	264	58	7
18	342	331	262	58	7
19	341	332	262	58	7
20	341	331	263	58	7
21	342	331	262	58	7
22	341	331	263	58	7
23	341	331	263	58	7
24	341	331	263	58	7

处于状态 i 的概率. 上述事实表明这个马尔可夫链在时刻 n 的时候所处状态的概率分布将收敛于一定的极限.

一般地，如果令马尔可夫链的状态矩阵为 $\boldsymbol{P}$, 在时刻 t 处于状态 i 的概率为 $\pi_n(i)$, 并令 $\boldsymbol{\pi}_n = (\pi_n(1), \pi_n(2)\cdots)$. 则仿式 (7.46),

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_{n-1}\boldsymbol{P} \quad (n \geq 1) \tag{7.47}$$

最后，反复利用式 (7.47)

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\boldsymbol{P}^n \tag{7.48}$$

(我们知道 n 步转移概率矩阵可由 $\boldsymbol{P}^n$ 给出，上式可由此直接给出). 当 n 趋近于无穷大的时候，如果假定 $\boldsymbol{\pi}_n$ 趋近于与 $\boldsymbol{\pi}_0$ 无关的极限分布 (用 $\boldsymbol{\pi}$ 记之), 则根据式 (7.47),

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P} \tag{7.49}$$

并称之为**平衡方程式**(equilibrium equation). 此外，式 (7.48) 的 $\boldsymbol{P}^n$ 收敛于一定的极限，令其极限为 $\boldsymbol{P}^{(\infty)}$, 则

$$\boldsymbol{\pi} = \boldsymbol{\pi}_0\boldsymbol{P}^{(\infty)} \tag{7.50}$$

最后，无论 $\boldsymbol{\pi}_0$ 取何值，为了使式 (7.50) 成立， $\boldsymbol{P}^{(\infty)}$ 同一列的所有元素应相等. 即

$$\boldsymbol{P}^{(\infty)} = \begin{bmatrix} v_1 & v_2 & \cdot & \cdot \\ v_1 & v_2 & \cdot & \cdot \\ v_1 & v_2 & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{7.51}$$

根据式 (7.50) 及式 (7.51), 我们可以推出 $\boldsymbol{P}^{(\infty)}$ 第 j 列的值 v_j 与 $\boldsymbol{\pi}$ 的第 j 项 $\pi(j)$ 相等. 即

$$\boldsymbol{\pi} = (v_1, v_2, \cdots) \tag{7.52}$$

换灯泡的例子中，以精确到小数点后 5 位数的精度去算 $P^2, P^4, P^8, P^{16}, P^{32}$，则我们清楚地看到将逼近式 (7.51)，如果在各列的取值再乘以 1000，则我们还可以看到它于表 7-2 最后一行的取值基本一致.

如果存在满足平衡方程式 (7.49) 的概率分布，则我们称 π 是这个马尔可夫链的 **平稳分布** (stationary distribution). 另外，$n \to \infty$ 的时候，如果 P^n 趋近于形如式 (7.51) 的极限形式，则称 $(v_1, v_2, \cdots)$ 是这个马尔可夫链的 **极限分布** (long-run distribution). 从上面的讨论我们清楚地看到，如果某个马尔可夫链存在极限分布，则这个极限分布也是它的平稳分布，但正如下面例子所述，即使存在平稳分布，却不一定存在极限分布.

例 7-11(周期性马尔可夫链) (periodic Markov chain) 下面考虑只有两个状态且转移概率矩阵为

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

的马尔可夫链. 这个系统的周期为 2, 且 $P^2 = I$ (单位矩阵), 如果 n 是奇数，则 $P^n = P$, 如果 n 是偶数，则 $P^n = I$, 即它不收敛于某一极限. 但 $\pi = (0.5, 0.5)$ 却是一个满足平衡方程式 (7.49) 的概率分布. 可见，这个马尔可夫链虽然不存在极限分布，但存在平稳分布.

下面，考察一下平稳分布及极限分布存在的条件.

首先，如果 π满足平衡方程式 (7.49), 则对于任意 n, 它还满足

$$\pi = \pi P^n \tag{7.53}$$

如果状态 j 是非常返的或是零常返的，则由定理 7-3 的推论 1 及定理 7-6, 对于任意的 i 都有 $P^{(n)}(i,j) \to 0$(当 $n \to \infty$ 的时候),

所以满足式 (7.53) 的 π 的第 j 元素 $\pi(j)$ 为 0. 也就是说，当马尔可夫链的所有状态都是非常返的或零常返的情况下，满足式 (7.53) 的 π 只能是 $\pi=0$. 因为这时不存在平稳分布，所以也不存在极限分布. 下面考虑存在正常返状态时的情况.

马尔可夫链的状态集合 S 由正常返的状态集合 S_1 及其它状态集合 S_2 构成，我们先把所有属于 S_1 的状态排在那些属于 S_2 的状态的前面，则得到如下转移概率矩阵 (因为正常返的等价类都是闭的，所以即使存在多个这样的等价类，S_1 仍是闭的).

$$\begin{array}{c} \\ S_1 \\ S_2 \end{array}\begin{array}{c} \begin{array}{cc} S_1 & S_2 \end{array} \\ \left[\begin{array}{c|c} \boldsymbol{P}_1 & \boldsymbol{0} \\ \hline \boldsymbol{Q} & \boldsymbol{R} \end{array}\right] \end{array} = \boldsymbol{P} \tag{7.54}$$

这里，$\boldsymbol{P}_1$ 表示属于 S_1 的状态之间的转移关系，且它自己构成一个转移概率矩阵. 在式 (7.54) 中我们把 $\boldsymbol{P}$ 分成了几个部分，相应地，我们把 π 也分成 $(\pi_1 \vdots \pi_2)$ 的形式. 则根据平衡方程式 (7.49), 有

$$\begin{aligned} \pi_1 &= \pi_1 \boldsymbol{P}_1 + \pi_2 \boldsymbol{Q} \\ \pi_2 &= \pi_2 \boldsymbol{R} \end{aligned}$$

这时，由于 $\pi_2 = 0$, 所以

$$\pi_1 = \pi_1 \boldsymbol{P}_1$$

从上面的讨论可以看出，当马尔可夫链中存在非常返状态的时候，只要考察除去这些状态以后所得到的系统的平衡方程式即可了. 所以此后，我们暂时只考察由正常返状态构成的马尔可夫链.

因为所有正常返的等价类都是闭的，所以当某个马尔可夫链中含有两个及两个以上等价类的时候，系统的转移概率矩阵

$$
\boldsymbol{P}=\left[\begin{array}{c:c:c}
\boldsymbol{P}_1 & 0 & 0\cdots \\ \hdashline
0 & \boldsymbol{P}_2 & 0\cdots \\ \hdashline
0 & 0 & \ddots \\
\vdots & \vdots &
\end{array}\right] \tag{7.55}
$$

这里，$\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots$ 表示每各等价类的转移概率矩阵．相应地，当我们令 $(\boldsymbol{\pi} = \boldsymbol{\pi}_1 \vdots \boldsymbol{\pi}_2 \vdots \cdots)$ 以后，平衡方程式 (7.49) 可依等价类写成

$$
\boldsymbol{\pi}_1 = \boldsymbol{\pi}_1 \boldsymbol{P}_1, \quad \boldsymbol{\pi}_2 = \boldsymbol{\pi}_2 \boldsymbol{P}_2, \cdots \tag{7.56}
$$

的形式．

如果假定式 (7.56) 的各方程式是相互独立的平衡方程式，且假定存在满足这些平衡方程式的平稳概率分布为 $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \cdots$, 那么当 $\omega_1, \omega_2, \cdots$ 为非负且其和为 1 的时候

$$
\boldsymbol{\pi}_\omega = (\omega_1 \boldsymbol{\pi}_1 \vdots \omega_2 \boldsymbol{\pi}_2 \vdots \cdots) \tag{7.57}
$$

成为转移概率矩阵 $\boldsymbol{P}$ 的平稳概率分布．

综上，某一马尔可夫链是否存在平稳概率分布的问题可以归结到由一个正常返的 (既约的) 等价类构成的马尔可夫链是否存在平稳分布这样一个问题．很显然，如果某一马尔可夫链具有两个及两个以上的等价类，则肯定不会存在极限分布．
这是因为

$$
\boldsymbol{P}^n=\left[\begin{array}{c:c:c}
\boldsymbol{P}_1^n & 0 & 0\cdots \\ \hdashline
0 & \boldsymbol{P}_2^n & 0\cdots \\ \hdashline
0 & 0 & \ddots \\
\vdots & \vdots &
\end{array}\right]
$$

而如果假定 $\boldsymbol{P}^n$ 逼近形如式 (7.51) 的极限 $\boldsymbol{P}^{(\infty)}$, 则 $\boldsymbol{P}^\infty$ 只能等于 0.

下面考虑只存在正常返状态的既约马尔可夫链，并令其转移概率矩阵为 $\boldsymbol{P}$. 因为对于任意两个状态 i 和 j, 都存在 $f(i,j)=1$, 故由定理 7-6

$$v(j) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P^{(l)}(i,j) = \frac{1}{m(j,j)} \tag{7.58}$$

定理 7-7 $\boldsymbol{v}=(v(1),v(2),\cdots)$ 是正常返的既约马尔可夫链的平稳分布，且此外不存在其它平稳分布.

证明 对于所有 j, $v(j)=1/m(j,j)>0$, 且

$$\begin{aligned}\sum_{j\in S} v(j) &= \sum_{j\in S} \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P^{(l)}(i,j)\\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \sum_{j\in S} P^{(l)}(i,j)\\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} 1\\ &=1\end{aligned}$$

所以，$\boldsymbol{v}=(v(1),v(2),\cdots)$ 是概率分布. 再由

$$\begin{aligned}\sum_{j\in S} v(j)P(j,k) &= \sum_{j\in S} \{\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P^{(l)}(i,j)\} P(j,k)\\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \sum_{j\in S} P^{(l)}(i,j)P(j,k)\\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} P^{(l+1)}(i,k)\\ &= \lim_{n\to\infty} \frac{1}{n+1} \{\sum_{l=1}^{n+1} P^{(l)}(i,k) - P(i,k)\}\\ &=v(k)\end{aligned}$$

v 是平衡方程式 (7.49) 的解，即 v 是平稳分布.

下面证明除 v 之外不存在其它平稳分布. 因为平稳分布 π 满足式 (7.53), 故在 $n=1,2,\cdots$ 的时候将它们相加平均，则

$$\pi = \pi \frac{1}{n} \sum_{l=1}^{n} P^{l}$$

这里，

$$\pi(j) = \sum_{i \in S} \pi(i) \frac{1}{n} \sum_{l=1}^{n} P^{(l)}(i,j)$$

在它两边取极限 $(n \to \infty)$, 则

$$\pi(j) = \sum_{i \in S} \pi(i) v(j) = v(j)$$

最后我们考虑正常返的既约马尔可夫链存在极限分布的条件. 正如前面我们所述，如果存在极限分布， P^n 将收敛于形如式 (7.51) 的极限. 这时根据式 (7.58), 对于任意 i,j 都有

$$\lim_{n \to \infty} P^{(n)}(i,j) = v(j) = 1/m(j,j) \tag{7.59}$$

且 $v = (v(1), v(2), \cdots)$ 是其唯一的极限分布. 我们已经举过一个例子说明周期性马尔可夫链中 P^n 不收敛于某一极限，这一般来讲是正确的. 而实际上，关于它我们还存在以下定理.

定理 7-8 令 i 和 j 为正常返的既约马尔可夫链的任意两个状态. 如果这个系统是非周期的，则

$$\lim_{n \to \infty} P^{(n)}(i,j) = 1/m(j,j) \tag{7.60}$$

如果这个系统的周期 $d \geq 2$, 则 $P^{(n)}(i,j)$ 不收敛.

但是，如果令 l_{ij} 为满足 $P^{(l)}(i,j)>0$ 的 l 的最小值，则

$$\lim_{n\to\infty} P^{(l+nd)}(i,j)=\begin{cases} d/m(j,j) & (l=l_{ij}) \\ 0 & \text{其它} \end{cases} \tag{7.61}$$

综上所述，我们得到如下关于极限分布的定理.

定理 7-9 马尔可夫链存在极限分布的充分必要条件为该系统是既约的，正常返的且是非周期的. 此时，平衡方程式的解 $\boldsymbol{\pi}$ 就是其极限分布，但此极限分布的第 j 元素 $\pi(j)$ 等于状态 j 的平均返回时间的倒数.

例 7-12 (选厂家；继续) 本章第一节例 7-1 的马尔可夫链是正常返的，且只有一个等价类. 所以它存在平稳分布及极限分布 $\boldsymbol{\pi}=(\pi(1),\pi(2),\pi(3),\pi(4))$, 且它满足平衡方程式 $\boldsymbol{\pi}=\boldsymbol{\pi}\boldsymbol{P}$, 即

$$\begin{aligned}
\pi(1)&=0.90\pi(1)+0.10\pi(2)+0.08\pi(3)+0.10\pi(4)\\
\pi(2)&=0.05\pi(1)+0.80\pi(2)+0.10\pi(3)+0.10\pi(4)\\
\pi(3)&=0.03\pi(1)+0.05\pi(2)+0.80\pi(3)+0.10\pi(4)\\
\pi(4)&=0.02\pi(1)+0.05\pi(2)+0.02\pi(3)+0.70\pi(4)
\end{aligned}$$

最后由 $\pi(1)+\pi(2)+\pi(3)+\pi(4)=1$, 我们可以求出

$$\boldsymbol{\pi}=(0.482,0.253,0.179,0.086)$$

可见，经过很长一段时间以后，厂家 A,B,C,D 所拥有的市场占有率将分别变为约 48%, 25%, 18% 和 9%(这个市场占有率是否正确，与是否在很长一段时间内转移概率矩阵保持恒定，即在时间上齐次，也就是说每个人购买啤酒的频率及每次购买量都相同

等条件有关).

有限马尔可夫链的情况 状态个数有限的马尔可夫链的平稳分布及极限分布与转移概率矩阵的特征值及特征向量有关. 下面我们讲一讲它们的关系. 一般地, 在实际问题中, 我们不去考虑特征值而通过上面所述平衡方程式去求平稳分布极限分布.

根据前面所述, 现在我们只需考虑系统既约的情况. 首先考察换灯泡例子的式 (7.45). 这里我们不用具体数值而用字母表示初始条件, 那么仿例 7-7 的随机游动问题, 式 (7.45) 可以通过差分方程的解法求出, 即令

$$v_n(j) = c_j\lambda^n \quad (\lambda \neq 0), \quad (1 \leq j \leq 5)$$

并把它代入式 (7.45), 然后在所有式子的两边除以 λ^{n-1}, 则

$$\begin{aligned}
c_1\lambda &= p_1c_1 + p_2c_2 + p_3c_3 + p_4c_4 + p_5c_5 \\
c_2\lambda &= q_1c_1 \\
c_3\lambda &= q_2c_2 \\
c_4\lambda &= q_3c_3 \\
c_5\lambda &= q_4c_4
\end{aligned}$$

上式还可用行向量 $\boldsymbol{c} = (c_1, c_2, \cdots, c_5)$ 及式 (7.44) 的转移概率矩阵 $\boldsymbol{P}$ 表示, 即

$$\lambda\boldsymbol{c} = \boldsymbol{c}\boldsymbol{P} \tag{7.62}$$

这里因为 $\boldsymbol{c}$ 是非零向量, 所以为了使式 (7.62) 成立, λ 应为特征方程式

$$|\lambda\boldsymbol{I} - \boldsymbol{P}| = 0 \tag{7.63}$$

的根 ($\boldsymbol{I}$ 为单位矩阵), 即 λ 应为 $\boldsymbol{P}$ 的特征值. 现在假定 $\boldsymbol{P}$ 的特征

值为 $\lambda_1, \lambda_2, \cdots, \lambda_5$, 且它们各不相同，则一般地，

$$v_n(j) = c_{j1}\lambda_1^n + c_{j2}\lambda_2^n + \cdots + c_{j5}\lambda_5^n \quad (1 \le j \le 5) \tag{7.64}$$

特别地，比如假定 λ_2 等于 λ_1, 则上式右边第二项可变为 $nc_{j2}\lambda_2^n$, 又假定 λ_5 也等于 λ_1 , 则式 (7.64) 右边第三项变为 $n^2c_{j3}\lambda_3^n$. 而无论什么情况，未定系数 $(c_{j1}, c_{j2}, \cdots, c_{j5})$ 都将由初始条件算出. 很显然，为了使 $v_n(j)$ 在 $n \to \infty$ 的时候不依赖于初始条件而收敛于一定极限，所有特征值的绝对值都应小于等于 1, 且绝对值等于 1 的特征值不能彼此相等. 但这些条件还是不够充分. 因为在特征值中除了 1 以外，还有形如

$$\lambda = e^{i\theta} = \cos\theta + i\sin\theta \quad (0 < \theta < 2\pi)$$

的绝对值为 1 的量，使得

$$\lambda^n = e^{ni\theta} = \cos n\theta + i\sin n\theta$$

在 $n \to \infty$ 的时候不收敛于某一极限值.

另一方面，我们还知道 $\boldsymbol{P}$ 中肯定存在特征值 $\lambda = 1$. 因为 $\boldsymbol{P}$ 的所有行元素之和为 1, 特征方程 (7.63) 的矩阵 $[\lambda \boldsymbol{I} - \boldsymbol{P}]$ 的所有行元素之和等于 $\lambda - 1$, 所以， $\lambda = 1$ 必为其特征值.

综上， $v_n(j)$ 收敛于某一极限，且它不依赖于初始条件的充要条件是，矩阵 $\boldsymbol{P}$的特征值 1 为单纯的，[1] 而其它特征值的绝对值都小于 1. 在上面的讨论中，我们没有用到形如式 (7.44) 的特殊的 $\boldsymbol{P}$, 所以它对于一般的马尔可夫过程都成立. 特别地，因为平衡方程式 (7.49) 相当于在式 (7.62) 中令 $\lambda = 1$, $\boldsymbol{c} = \boldsymbol{\pi}$, 故其解 $\boldsymbol{\pi}$ 实际上是 $\boldsymbol{P}$ 的特征值 1 所对应的特征向量. [2]

---

[1] 即它是特征方程 (7.63) 的单根.

[2] 解析学中，把满足 $\boldsymbol{P}\boldsymbol{\pi} = \lambda\boldsymbol{\pi}$ 的列向量称做特征向量. 而在这里因为 $\boldsymbol{\pi}$ 是行向量，所以应该左乘 $\boldsymbol{P}$.

转移概率矩阵特征值的性质非常重要. 下面我们介绍代数学中非常有名的佩龙-弗洛贝尼斯定理 (Perron-Frobenius theorem). 这是有关所有元素为非负的矩阵 (行元素之和不一定为 1) 特征值的定理. 现在, 我们结合本书来叙述这个定理.

定理 7-10 假定 $\boldsymbol{P}$ 为既约有限马尔可夫链的转移概率矩阵, 则

(1) $\boldsymbol{P}$ 具有特征值 1, 且它是单纯的. 另外所有其它特征值的绝对值即小于等于 1.

(2) $\boldsymbol{P}$ 的对应于特征值 1 的特征向量, 即方程

$$\pi \boldsymbol{P} = \pi$$

的解的所有元素可为正 (换一句话讲, 在 π 的所有元素之和为 1 的条件下解上述方程式, 则 π 的所有元素必为正).

(3) 假定绝对值为 1 的 $\boldsymbol{P}$ 的特征值的个数为 m, 则这些特征值是 1 的 m 次根, 即它们等于 [1]

$$e^{2\pi ik/m} \qquad (k = 0, 1, \cdots, m-1)$$

(4) 除 $\lambda = 1$ 以外的 $\boldsymbol{P}$ 的所有特征值的绝对值小于 1 的 [2] 充要条件是, 存在自然数 k, 使得 $\boldsymbol{P}^k$ 的所有元素为正.

这个定理说明了非周期的有限既约马尔可夫链是否存在极限分布将取决于转移概率矩阵 $\boldsymbol{P}$ 是否是原始的. 但为了通过 (4) 判定是否存在极限分布, 我们还需要知道 $\boldsymbol{P}$ 原始的时候, 使得 $\boldsymbol{P}^k$ 的所有元素为正的自然数 k 的最小值的上界. $\boldsymbol{P}$ 的阶数为 l, 那

---

[1] 我们把包含 $\lambda = 1$ 的 m 个绝对值为 1 的特征值表示在一个复平面上, 则这些点就会等间隔地排列在以原点为中心的单位圆上.

[2] 这时, 我们称 $\boldsymbol{P}$ 是原始的 (primitive).

么我们知道 l^2-2l+2 就是这样的一个上界. 现在假定, 于是, 当 $k=1,2,\cdots$ 的时候, 我们依次考察 $\boldsymbol{P}^k$ 的符号, 如果在某一处所有元素变为正, 则可以判定 $\boldsymbol{P}$ 是原始的. 反过来, 如果直到 $k=l^2-2l+2$, 我们还找不到这样一个 $\boldsymbol{P}^k$, 则可以判定 $\boldsymbol{P}$ 不是原始的.

习 题 7

1. 根据连通性, 试将下面马尔可夫链的转移概率矩阵按等价类进行划分, 并考察各等价类的周期 d 及常返性.

(a)

$$\begin{bmatrix} 0.2 & 0 & 0.8 & 0 & 0 \\ 0.4 & 0.3 & 0.3 & 0 & 0 \\ 0.7 & 0 & 0 & 0.3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0 & 0.1 & 0.4 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 0.5 & 0 & 0 & 0 & 0.5 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(c)

$$\begin{bmatrix} 0 & 0.2 & 0.3 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 & 0 & 0 & 0 \\ 0.4 & 0 & 0 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.3 & 0 & 0 & 0 & 0 & 0.7 & 0 \end{bmatrix}$$

(d)

$$\begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 & 0.2 & 0.3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0.7 & 0 & 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

2. 就第一题 (a) 的情况，回答以下问题.

(1) 试将状态按非常返的状态状态集合 T 和常返的状态集合 $S-T$ 分类.

(2) 试将转移概率矩阵写成形如式 (7.38) 的形式.

(3) 试求基础矩阵 $\boldsymbol{N}$.

(4) 试求从属于 T 的各状态 i 出发到离开 T 为止所需转移次数的数学期望 $m(i)$.

3. 考察 $S=\{0,1,2,\cdots,K\}$ 上的随机游动. 假定两端 $0,K$ 为吸收壁，从中间的某个状态向右移动一步的概率为 p, 向左移动一步的概率为 $q=1-p, p\neq q$.

(1) 试求 $f(i,0)$ 及 $f(i,K)$, $(1\leq i\leq K-1)$.

(2) 试求从状态 i 出发的系统到达吸收壁所需转移次数的数学期望 $m(i)$.

4. 考察状态数有限，非常返的状态集合 T 及常返的状态集合 $S-T$ 均为非空的马尔可夫链. 假定每个常返的状态都是闭的 (单一的闭的状态称做**吸收状态** (absorbing state)). 利用本章第 3 节所述记号，回答以下问题.

(1) 设从状态 $i\in T$ 出发的系统被 $j\in S-T$ 吸收的概率 $f(i,j)$ 构成矩阵 $\boldsymbol{F}$. 试证 $\boldsymbol{F}=\boldsymbol{NR}$.

(2) 设系统从状态 $i \in T$ 出发被某一属于 $S-T$ 的状态吸收为止所需时间的二次 (原点) 矩为 $m^{(2)}(i)$, 又设由 $m^{(2)}(i)(i \in T)$ 构成的列向量为 $\boldsymbol{m}^{(2)}$, 试证

$$\boldsymbol{m}^2 = \boldsymbol{N}(2\boldsymbol{m} - \mathbf{1})$$

(3) 试求例 7-2 及例 7-8 围棋赛的 F. 另外, 试求 $p = q = r = 1/2$ 时的 $\boldsymbol{m}^{(2)}$.

5. 按下面的顺序, 试证若状态 i, j 相通, 则周期 $d(i)$ 等于 $d(j)$ ([定理 7-2]).

(1) 令 $P^{(l)}(i,j) > 0, P^{(m)}(j,j) > 0, P^{(n)}(j,i) > 0$.

(2) 证明 $P^{(2m)}(j,j) > 0$.

(3) 证明 $P^{(l+m+n)}(i,i) > 0, P^{(l+2m+n)}(i,i) > 0$.

(4) 证明 $d(i)$ 能被 $(l+2m+n)-(l+m+n)=m$ 整除.

(5) 证明 $d(i)$ 能被 $d(j)$ 整除.

(6) 证明 $d(i) = d(j)$.

6. 就第 1 题 (b) 的情况, 判定是否存在平稳分布及极限分布. 如果存在, 试求之.

7. **埃伦菲斯特模型** (Ehrenfest's model) 瓶中装有 a 个红球和白球. 现从瓶中拿出一个球, 并把另外一个与此球颜色相反的球放到瓶子中去. 这种操作, 无限反复进行下去. 假定经过 n 次操作后瓶子中红球的个数为 X_n, 那么 $\{X_n; n \geq 0\}$ 是一个以 $\{0, 1, 2, \cdots, a\}$ 为状态空间的马尔可夫链.

(1) 试求出此系统的转移概率矩阵.

(2) 问是否存在极限分布及平稳分布. 如果存在, 试求之.

8. 若转移概率矩阵任意列的元素之和等于 1, 则此矩阵被称做**二重概率矩阵** (doubly stochastic). 如果某一非周期的既约马尔可夫链的状态个数 a 有限, 且其转移概率矩阵 $\boldsymbol{P}$ 是二重概率, 试

证存在极限分布 π, 且

$$\pi(1) = \pi(2) = \cdots = \pi(a) = 1/a$$

附　　录

A　拉普拉斯变换

拉普拉斯变换的详细内容及各种公式，可参照江泽洋的《傅里叶解析》和森口繁一等的《数学公式 II— 级数. 傅里叶解析》(岩波书店). 这里，我们仅以本书出现的内容为中心，简单介绍它的一些基本性质及有关公式.

假定函数 $f(t)$ 满足以下性质：

1. $t<0$ 的时候， $f(t)=0$.

2. 从右边趋近于 $t=0$ 的时候，存在极限值 $f(+0)$.

3. $t>0$ 的时候，它是分段连续的，且各间断点处的左右极限都存在.

4. 对于某一实数 z, 积分

$$\tilde{f}(z)=\int_0^{\infty} f(t)e^{-zt}\mathrm{d}t$$

收敛.

这时，称 $\tilde{f}(z)$ 为 $f(t)$ 的拉普拉斯变换 (拉普拉斯变换的参考书中，大多用 s 而不用 z. 可是在本书中，我们用 z 表示 s. 注意这里的 z 并不是复数). 用 $\tilde{f}(z)$ 表示 $f(t)$ 的拉普拉斯变换的时候，我们可把它记做 $f(t)\to\tilde{f}(z)$.

拉普拉斯变换的性质

1. 线性性质

$$Af(t)+Bg(t)\rightarrow A\widetilde{f}(z)+B\widetilde{g}(z),\quad (A,B\text{为定常数})$$

2. 微分 → 乘法

$$\frac{\mathrm{d}}{\mathrm{d}t}f(x)\rightarrow z\widetilde{f}(z)-f(+0)$$

3. 积分 → 除法

$$\int_0^t f(u)\mathrm{d}u\rightarrow\frac{1}{z}\widetilde{f}(z)$$

4. 乘法 → 微分

$$tf(t)\rightarrow-\frac{\mathrm{d}}{\mathrm{d}z}\widetilde{f}(z)$$

$$t^n f(t)\rightarrow(-\frac{\mathrm{d}}{\mathrm{d}z})^n\widetilde{f}(z)$$

5. 除法 → 积分

$$\frac{1}{t}f(t)\rightarrow\int_z^\infty\widetilde{f}\,u\mathrm{d}u$$

6. 线性变换

$$f(at)\rightarrow\frac{1}{a}\widetilde{f}(\frac{z}{a})\qquad(a>0)$$

$$f(t-b)\rightarrow e^{-bt}\widetilde{f}(z)\qquad(b>0)$$

$$e^{ct}f(t)\rightarrow\widetilde{f}(z-c)$$

7. 卷积 → 积

$$\int_0^t f(t-u)g(u)\mathrm{d}u\rightarrow\widetilde{f}(z)\,\widetilde{g}(z)$$

简单函数的拉普拉斯变换

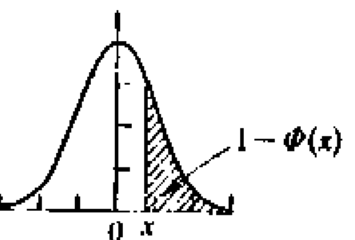
$f(t)$	$\tilde{f}(z)$	$f(t)$	$\tilde{f}(z)$
1	$\frac{1}{z}$	$\frac{t^{n-1}e^{-at}}{(n-1)!}$	$\frac{1}{(z+a)^n}$
t	$\frac{1}{z^2}$	$\sin bt$	$\frac{b}{z^2+b^2}$
$\frac{t^{n-1}}{(n-1)!}$	$\frac{1}{z^n}$	$\cos bt$	$\frac{z}{z^2+b^2}$
e^{-at}	$\frac{1}{z+a}$	$e^{-at}\sin bt$	$\frac{b}{(z+a)^2+b^2}$
te^{-at}	$\frac{1}{(z+a)^2}$	$e^{-at}\cos bt$	$\frac{z+a}{(z+a)^2+b^2}$

n 为自然数， a 为正数， b 为任意实数

B　标准正态分布的上侧概率 $1-\Phi(x)$

下面这个表，对于 $0.00 \sim 3.00$ 的 x，以 0.01 的间隔给出了服从标准正态分布 $N(0,1)$ 的随机变量 X 大于等于 x 的概率 $P(X \geq x)$. 比如，$P(X \geq 1.23)$ 的值 0.1093 可从左侧 1.2* 行，上侧 $* = 3$ 列的位置查出.

B 标准正态分布的上侧概率 $1-\Phi(x)$



x	*=0	1	2	3	4	5	6	7	8	9
0.0*	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1*	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2*	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3*	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4*	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5*	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6*	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7*	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8*	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9*	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0*	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1*	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2*	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3*	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4*	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5*	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6*	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7*	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367

1.8*	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9*	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0*	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1*	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2*	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3*	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4*	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5*	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6*	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7*	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8*	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9*	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0*	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

习 题 答 案

习 题 1

1. (1) $\Omega=\{(H,H,H),(H,H,T),(H,T,H),(T,H,H),$
$(H,T,T),(T,H,T),(T,T,H),(T,T,T)\}$
(2) $A=\{(H,H,H),(H,H,T),(H,T,H),(H,T,T)\}$,
$B\cup C=\{(H,H,H),(H,H,T),(H,T,H),(T,H,H),(T,H,T),$
$(T,T,H)\}$, $B\cap C=\{(H,H,H),(T,H,H)\}$,
$A^c\cup B=\{(T,H,H),(T,H,T)\}$.
$(A\cup B^c)\cap C=\{(H,H,H),(H,T,H),(T,T,H)\}$
(3) 1/2, 3/4, 1/4, 1/4, 3/8.

2. (1) $A\cup B\cup C$ (2) $A^c\cap B^c\cap C^c$ (3) $(A^c\cap B\cap C)\cup(A\cap B^c\cap C)\cup(A\cap B\cap C^c)$ (4) $(A\cap B)\cup(B\cap C)\cup(C\cap A)$ (5) $((A\cap B^c)\cup(A^c\cap B))\cap C^c$

3. $P=1-p^n,(n\geq 2)$

4. 假定跳一次成功的概率为 $p=0.3$, $q=1-p$, 则 $p+qp+q^2p=(1+0.7+0.7^2)\times 0.3=0.657$

5. 略.

6. (1) 如果父母的血型组合为 O×B, 则从 B 型父母那里得到遗传因子 O 的时候, 孩子的血型为 O 型, 否则为 B 型. B 型父母遗传因子型为 BB 的概率 P(BB)=3/22, BO 的概率 P(BO)=19/22, 故由全概率公式, 孩子血型为 O 的概率 $P(\text{O})=P(\text{O}|\text{BB})P(\text{BB})+P(\text{O}|\text{BO})P(\text{BO})=1\times(3/22)+(1/2)\times(19/22)\approx 0.57$, 孩子血型为 B 型的概

率 $P(\mathrm{B})=1-0.57=0.43$. 对于父母血型的其它组合，我们可依此类推.

(2) 根据贝叶斯公式，

$$P(\mathrm{AO}|\mathrm{A})=\frac{P(\mathrm{A}|\mathrm{AO})P(\mathrm{AO})}{P(\mathrm{A}|\mathrm{AO})P(\mathrm{AO})+P(\mathrm{A}|\mathrm{AA})P(\mathrm{AA})}$$
$$=\frac{0.5\times0.31}{0.5\times0.31+1\times0.08}\approx0.66$$

(3) 父母的遗传因子型组合与孩子的遗传因子型关系如下表所示.

	BB	BO
AA	AB	AB, AO
AO	AB, BO	AB, AO, BO, OO

如果框内孩子的遗传因子型有两个及两个以上，则它们之中任意一个以等概率出现. 故由贝叶斯公式

$$P(\mathrm{AO}|\mathrm{AB})=\frac{0.5\times0.31\times0.03+0.25\times0.31\times0.19}{0.08\times0.03+0.5\times0.08\times0.19+0.5\times0.31\times0.03+0.25\times0.31\times0.19}$$
$$\fallingdotseq0.66$$

$$P(\mathrm{BB}|\mathrm{AB})=\frac{0.08\times0.03+0.5\times0.31\times0.03}{0.08\times0.03+0.5\times0.08\times0.19+0.5\times0.31\times0.03+0.25\times0.31\times0.19}$$
$$\fallingdotseq0.24$$

习 题 2

1. 根据泊松分布的再生性，我们知道 $X+Y$ 服从泊松分布

$\text{Po}(\lambda_1+\lambda_2)$，故

$$\begin{aligned}P(X=k|X+Y=l)&=P(X=k,X+Y=l)/P(X+Y=l)\\&=P(X=k,Y=l-k)/P(X+Y=l)\\&=e^{-\lambda_1}\frac{\lambda_1^k}{k!}\cdot e^{-\lambda_2}\frac{\lambda_2^{l-k}}{(l-k)!}/e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^l}{l!}\\&=\frac{l!}{k!(l-k)!}(\frac{\lambda_1}{\lambda_1+\lambda_2})^k(\frac{\lambda_2}{\lambda_1+\lambda_2})^{l-k}\end{aligned}$$

即 X 的条件分布服从二项分布 $\text{B}(l;\lambda_1/(\lambda_1+\lambda_2))$.

2. 令 k 为任意非负整数，则

$$\begin{aligned}&P(X=k)\\&=\sum_{l=0}^{\infty}P(X=k|Z=l)P(Z=l)=\sum_{l=k}^{\infty}\binom{l}{k}p^k(1-p)^{l-k}\cdot e^{-\lambda}\frac{\lambda^l}{l!}\\&=e^{-\lambda}\frac{(\lambda p)^k}{k!}\sum_{l=k}^{\infty}\frac{\{\lambda(1-p)\}^{l-k}}{(l-k)!}=e^{-\lambda}\frac{(\lambda p)^k}{k!}e^{\lambda(1-p)}=e^{-\lambda p}\frac{(\lambda p)^k}{k!}\end{aligned}$$

故 X 服从泊松分布 $\text{Po}(\lambda p)$.

3. (1) 令 k 为非负整数，$q=1-p$, 则

$$\begin{aligned}P(X\geq l+k|X\geq l)&=P(X\geq l+k)/P(X\geq l)\\&=\sum_{m=l+k}^{\infty}q^mp/\sum_{m=l}^{\infty}q^mp=q^{l+k}/q^l=q^k\end{aligned}$$

(2) 令 x 为非负实数，则

$$\begin{aligned}P(Y\geq z+x|Y\geq z)&=P(Y\geq z+x)/P(Y\geq z)\\&=e^{-\alpha(z+x)}/e^{-\alpha z}\\&=e^{-\alpha x}\end{aligned}$$

很显然，在这二个分布之中，即使知道随机变量的值在某一确定值之上，我们却不知道它究竟有多大.

4. 令 $g(l)=P(x\geq l)$，则 $g(n+m)=g(n)g(m)$. 又令 $m=1$，则 $g(n+1)=g(1)g(n)$，即 $\{g(n)\}$ 为等比数列，故它可以表示成 $g(n)=aq^n$ 的形式。因为 $g(0)=P(x\geq 0)=1$，故 a 为 1，于是 $P(x=k)=P(x\geq k)-P(x\geq k+1)=q^k-q^{k+1}=pq^k$，$X$ 服从几何分布．

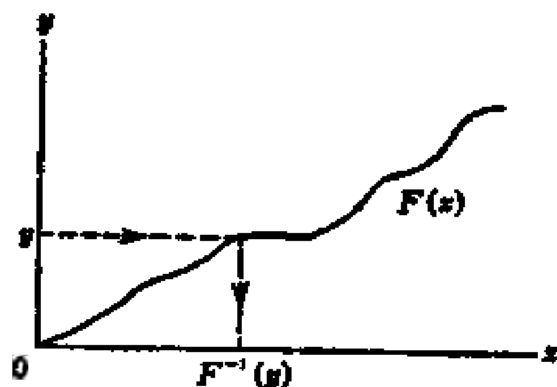
5. Y 可表示成 $Y=(-\log X_1)+\cdots+(-\log X_n)$ 的形式，而 $-\log X_i$ 服从指数分布 Ex(1)，故 Y 服从伽玛分布 G(1,n).

6. $P(Y\leq y)=P(Y_1\leq y,\cdots,Y_n\leq y)=P(Y_1\leq y)\cdots P(Y_n\leq y)=\{F(y)\}^n$.

$P(Z\leq z)=1-P(Z>z)=1-P(X_1>z,\cdots,X_n>z)$

$=1-P(X_1>z)\cdots P(X_n>z)=1-\{1-F(z)\}^n$.

7. 如图所示，定义 F 的逆函数 F^{-1} 为 $F^{-1}(y)=\inf\{x|F(x)=y\}$，



$$P(Y\leq y)=P(F(x)\leq y)$$
$$=P(X\leq F^{-1}(y))=F(F^{-1}(y))=y$$

即 Y 服从 U(0,1) 分布．

8. $P(U\leq u)=\int\int_{x+y\leq u}f(x,y)\mathrm{d}x\mathrm{d}y$

令 $t=x+y$，并对上式关于 x 及 t 进行积分变换，则 $P(U\leq u)=\int_0^u\mathrm{d}t\int_0^t f(x,t-x)\mathrm{d}x$，再对 u 求导，则 $f_U(u)=\int_0^u f(x,u-x)\mathrm{d}x$. 同理对于 V,W,Z，令其中的一个积分变量为 $x-y,xy,y/x$，则可得到以

下结果.

$$f_V(v)=\int_v^{\infty} f(x,x-v)\mathrm{d}x,\quad f_W(w)=\int_0^{\infty}\frac{1}{x}f(x,w/x)\mathrm{d}x$$

$$f_Z(z)=\int_0^{\infty} xf(x,zx)\mathrm{d}x.$$

9. 因为

$$\begin{aligned}Q&=\frac{1}{1-\rho^2}\{(\frac{x-\mu_1}{\sigma_1})^2-2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})+(\frac{y-\mu_2}{\sigma_2})^2\}\\&=\frac{1}{1-\rho^2}(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1})^2+(\frac{x-\mu_1}{\sigma_1})^2\equiv Q_1+(\frac{x-\mu_1}{\sigma_1})^2\end{aligned}$$

故

$$\begin{aligned}f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y&=\frac{1}{\sqrt{2\pi}\sigma_1}\exp[-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2]\\&\times\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp[-\frac{1}{2}Q_1]\mathrm{d}y\end{aligned}$$

固定 x, 则最后一个积分的被积函数相当于正态分布的概率密度函数, 即其积分值为 1, 所以

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp[-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2$$

即 X 的边缘分布服从 $\mathrm{N}(u_1,\sigma_1^2)$, 同理, Y 的边缘分布服从 $\mathrm{N}(u_2,\sigma_2^2)$.

10. 因为 $P(V\le v)=P(V^2\le v^2)$, $W=V^2=X_1^2+X_2^2+X_3^2$ 服从伽玛分布 $\mathrm{G}(\frac{1}{2\sigma^2},\frac{3}{2})$, 故

$$P(V\le v)=\int_0^{v^2}\frac{1}{\Gamma(3/2)}(\frac{1}{2\sigma^2})^{3/2}w^{3/2-1}\exp[-\frac{w}{2\sigma^2}]\mathrm{d}w$$

在两边关于 v 求导, 则 V 分布的概率密度函数 $f(v)$ 为

$$\begin{aligned}f(v)&=\frac{\mathrm{d}}{\mathrm{d}v}(v^2)\cdot\frac{1}{\Gamma(3/2)}(\frac{1}{2\sigma^2})^{3/2}(v^2)^{3/2-1}\exp[-\frac{v^2}{2\sigma^2}]\\&=\frac{\sqrt{2/\pi}}{\sigma^3}v^2\exp[-\frac{v^2}{2\sigma^2}]\quad(v\ge 0)\end{aligned}$$

11. 略.

12. 指数分布 Ex(α)

习题 3

1. 令 $g(X)$ 的取值为 $u_0, u_1, \cdots$, 并令 $g(X)$ 的概率分布函数为 $h(u)$, 则 $h(u_j) = P(g(X) = u_j) \quad (j = 0, 1, 2, \cdots)$. 对于任意固定的 j, 令满足 $g(v_k) = u_j$ 的 k 的集合为 $K(j)$, 则 $h(u_j) = \sum\limits_{k \in K(j)} f(v_k)$, 故

$$E(g(X)) = \sum_j u_j h(u_j) = \sum_j u_j \sum_{k \in K(j)} f(v_k)$$
$$= \sum_j \sum_{k \in K(j)} g(v_k) f(v_k) = \sum_k g(v_k) f(v_k).$$

2. 令 $q = 1-p$, 则 $E(Y) = \sum_{k=0}^{l} kq^k p + l \sum_{k=l+1}^{\infty} q^k p = (q/p)(1-q^l)$.

3. $E[(X-c)^2] = E[X^2 - 2cX + c^2] = c^2 - 2cE(X) + E(X^2)$ 为 c 的二次式, 即当 $c = E(X)$ 的时候其值最小.

4. 设相互独立地服从几何分布 Ge(p) 的随机变量 $X_1, \cdots X_r$ 之和为 Y, 则 Y 服从 NB($r;p$) 分布, 即 $E(Y) = r \cdot E(X_1) = rq/p$, $\mathrm{Var}(Y) = r \cdot \mathrm{Var}(X_1) = rq/p^2$.

5. $E(Z) = E(X) - \rho E(Y) = 0$, $\mathrm{Var}(Z) = \mathrm{Var}(X) + \rho^2 \mathrm{Var}(Y) - 2\rho \mathrm{Cov}(X, Y) = 1 + \rho^2 - 2\rho^2 = 1 - \rho^2$, $\mathrm{Cov}(Z, Y) = \mathrm{Cov}(X, Y) - \rho \mathrm{Var}(Y) = \rho - \rho = 0$.

6. 因为对于所有实数 t, $E\{(tX+Y)^2\} = E(X^2)t^2 + 2E(XY)t + E(Y^2) \geq 0$, 故 $\{E(XY)\}^2 - E(X^2)E(Y^2) \leq 0$. 等号成立的时候, 判别式的值为 0, 即存在实数 t, 使得 $E\{(tX+Y)^2\} = 0$, 此时 $tX + Y = 0$, 即 X 与 Y 成比例.

7. 令 X_1, X_2 的均值为 μ_1, μ_2, 并令 $X = X_2 - \mu_2$, $Y = X_1 - \mu_1$, 然后将它们代入到上一题的不等式中, 当 $X_2 - \mu_1$ 与 $X_1 - \mu_2$ 成比例, 且比例常数为正 (负) 的时候, 相关系数为 $1(-1)$.

8. 由定理 3-4

$$E(S_N)=\sum_{n=0}^{\infty}E(S_N|N=n)P(N=n)=\sum_{n=0}^{\infty}n\mu P(N=n)=E(N)\mu,$$

$$E(S_N^2)=\sum_{n=0}^{\infty}E(S_N^2|N=n)P(N=n)$$

$$=\sum_{n=0}^{\infty}[\mathrm{Var}(S_N|N=n)+\{E(S_N|N=n)\}^2]P(N=n)$$

$$=\sum_{n=0}^{\infty}[n\delta^2+(n\mu)^2]P(N=n)=E(N)\delta^2+E(N^2)\mu^2$$

故 $\mathrm{Var}(S_N)=E(S_N^2)-\{E(S_N)\}^2=E(N)\delta^2+$ $[E(N^2)-\{E(N)\}^2]\mu^2=E(N)\delta^2+\mathrm{Var}(N)\mu^2.$

9. (1)

$$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}\{(X_i-\mu)-(\overline{X}-\mu)\}^2$$

$$=\sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)+n(\overline{X}-\mu)^2$$

而 $\sum_{i=1}^{n}(X_i-\mu)=n(\overline{X}-\mu)$, 故 $\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2$.

(2) $E[\sum_{i=1}^{n}(X_i-\overline{X})^2]=E[\sum_{i=1}^{n}(X_i-\mu)^2]-E[n(\overline{X}-\mu)^2]=n\sigma^2-nE[(\overline{X}-\mu)^2]$. 这里，因为分布的均值为 μ，方差为 σ^2/n，故 $E[(\overline{X}-\mu)^2]=\sigma^2/n$, $E[\sum_{i=1}^{n}(X_i-\overline{X})^2]=(n-1)\sigma^2$

10.

	$P(\lvert Y\rvert\geq 1)$	$P(\lvert Y\rvert\geq 2)$	$P(\lvert Y\rvert\geq 3)$
上界评估	1.000	0.250	0.111
Ex(1)	0.135	0.050	0.018
$N(\mu,\sigma^2)$	0.318	0.046	0.002

很显然，针对所有分布的上界评估值与某一特定分布的概率值之间有很大的偏差.

习题 4

1. (1) $p/(1-qz)$ (2) 略. (3) $G(z;r)=p^r/(1-qz)^r$ (4) 因为 $G(z;r_2)=G(z;r_1+r_2)$, 故 (关于 r) 具有再生性.
(5) $p^r/(1-qz)^r=(1-\lambda/r)^r/\{1-(\lambda/r)z\}^r\to e^{\lambda}/e^{-1}=e^{\lambda(z-1)}$

2. (1) $\Phi(\frac{12.5-10}{\sqrt{10}})\approx 0.785$ (2) $1-\Phi(\frac{99.5-80}{\sqrt{400}})\approx 0.165$ (3) $1-\Phi(\frac{50-40}{\sqrt{40}})\approx 0.057$. 这些分布都具有再生性，且它们是相互独立地服从同一分布的多个随机变量之和的分布，故可利用中心极限定理.

3. (1) $(1-z)Q(z)=q_0+\sum_{k=1}^{\infty}(q_k-q_{k-1})z^k$
$=1-f(0)+\sum_{k=1}^{\infty}[-f(k)]z^k=1-G(z)$.

(2) 由式 (4.1), $Q(1)=G'(1)$. 在 $(1-z)Q(z)=1-G(z)$ 的两边连续二次求导，并令 $z=1$, 则 $2Q'(1)=G''(1)$, 它和 (4.8),(4.9)，我们得到 (4.45).

4. (1) 第 1 次掷骰子的时候，分 "1" 出现和 "1" 没有出现两种情况去考虑.

(2) 在式 (4.46) 的两边乘以 z^k, 然后由 $k=1,2,\cdots\infty$, 可以得到 $A(z)=\frac{1-(5/6)z}{[1-(2/3)z](1-z)}$.

(3) 对 $A(z)$ 进行级数展开，则

$$A(z)=\frac{1}{z}[\frac{1}{1-z}+\frac{1}{1-(2/3)z}]=\frac{1}{z}\sum_{k=0}^{\infty}[1+(\frac{2}{3})^k]z^k$$

即 $a_k=\frac{1}{2}[1+(\frac{2}{3})^k]$

5. (1) 如果在买第 n 盒糕点的时候，收集到 r 种彩券，则在买第 $r-1,r,r+1,\cdots,n-1$ 盒糕点的时候，必须收集到 $r-1$ 种彩券，现在假定在买第 $n-k$ 盒糕点的时候收集到第 $r-1$ 种彩券，则在买第 n 盒糕

点的时候收集到第 r 种彩券的(条件)概率为 $(\frac{r-1}{N})^{k-1}\times(1-\frac{r-1}{N})$. 故由全概率公式，我们得到式 (4.47).

(2) 在式 (4.47) 中用 $n+1$ 替换 n, 并将此式与式 (4.47) 进行比较即可.

(3) 在式 (4.48) 的两边乘以 z^{n+1}, 然后令 $n=0,1,\cdots,\infty$, 即可得到结果.

(4) 反复利用式 (4.49)，并利用 $p(z;1)=z$.

(5) 将式 (4.50) 写成如下形式：

$$P(z;r)=z^r\prod_{j=1}^{r-1}\frac{1-j/N}{1-(j/N)z}$$

这里，z^r 为取 r 值的概率分布的母函数，$(1-j/N)/[1-(j/N)z]$ 为几何分布 $\mathrm{Ge}(1-j/N)$ 的概率母函数.

6. 考虑收集到 r 种指定彩券中的任意一种彩券时所买的糕点箱的盒数 (k),

$$p(n;r)=\sum_{k=1}^{n-r+1}(1-\frac{r}{N})^{k-1}\frac{r}{N}P(n-k;r-1)$$

由于母函数满足 $P_r(z)=\frac{r}{N}zP_{r-1}(z)+(1-\frac{r}{N})zP_r(z)$, 故 $P_r(z)=z\prod_{j=2}^{r}\frac{jz}{N-(N-j)z}$.

7.

$$\begin{aligned}M_S(\theta)=E[\exp(\theta S_N)]&=\sum_{n=0}^{\infty}E[\exp(\theta S_N)|N=n]P(N=n)\\&=\sum_{n=0}^{\infty}\{M_X(\theta)\}^nP(N=n)\\&=E[\{M_X(\theta)\}^N]=G_N(M_X(\theta)).\end{aligned}$$

对 θ 求导，则

$$M_X'(\theta)=G_N'(M_X(\theta))M_X'(\theta)$$
$$M_S''(\theta)=G_N''(M_X(\theta))\{M_X(\theta)\}^2+G_N'(M_X(\theta))M_X''(\theta).$$

这里，令 $\theta=0$, 并注意到 $M_X(0)=1$, 则

$$\begin{aligned}
E(S_N) &= E(N)E(X)\\
E(S_N^2) &= E[N(N-1)]\{E(X)\}^2+E(N)E(X^2)\\
&= E(N^2)\{E(X)\}^2+E(N)\mathrm{Var}(X)
\end{aligned}$$

于是 $\mathrm{Var}(S_N)=E(S_N^2)-\{E(S_N)\}^2=\mathrm{Var}(N)\{E(X)\}^2$ $+E(N)\mathrm{Var}(X)$.

8. 令 $\overline{X}$ 的特征函数为 $\phi\overline{X}(t)$, 则

$$\begin{aligned}
\phi\overline{X}(t) &= E[\exp\{i\frac{X_1+X_2+\cdots+X_n}{n}t\}]\\
&= \prod_{k=1}^{n}E[\exp\{iX_k\frac{t}{n}\}]=\prod_{k=1}^{n}\varphi(\frac{t}{n})\\
&= [\exp\{i\mu\frac{t}{n}-\alpha|\frac{t}{n}|\}]^n=\exp[i\mu t-\alpha|t|]=\varphi(t).
\end{aligned}$$

即 $\overline{X}$ 与 X_1 等服从同一分布.

9. 令 $Y_1,Y_2,\cdots,Y_n$ 的特征函数为 $\varphi(t_1,t_2,\cdots,t_n)$, 则

$$\begin{aligned}
&\varphi(t_1,t_2,\ldots,t_n)\\
&= E[\exp\{i\sum_{k=1}^{n}t_kY_k\}]=E[\exp\{i\sum_{k=1}^{n}t_k\sum_{j=1}^{n}a_{kj}X_j]\\
&= E[\exp\{i\sum_{j=1}^{n}(\sum_{k=1}^{n}t_ka_{kj})X_j\}]=\prod_{j=1}^{n}E[\exp\{i(\sum_{k=1}^{n}t_ka_{kj})X_j\}]\\
&= \prod_{j=1}^{n}\exp\{-\frac{1}{2}(\sum_{k=1}^{n}t_ka_{kj})^2\}=\exp\{-\frac{1}{2}\sum_{j=1}^{n}(\sum_{k=1}^{n}t_ka_{kj})^2\}
\end{aligned}$$

这里，

$$\begin{aligned}
\sum_{j=1}^{n}(\sum_{k-1}^{n}t_ka_{kj})^2 &= \sum_{j-1}^{n}(\sum_{k-1}^{n}t_ka_{kj})(\sum_{l=1}^{n}t_la_{lj})\\
&= \sum_{k=1}^{n}\sum_{l=1}^{n}t_kt_l\sum_{j=1}^{n}a_{kj}a_{lj}.
\end{aligned}$$

再由正交条件，它等于 $\sum_{k=1}^{n} t_k^2$，故

$$\varphi(t_1,\cdots,t_n)=\exp\{-\frac{1}{2}\sum_{k=1}^{n}t_k^2\}=\prod_{k=1}^{n}\exp\{-\frac{1}{2}t_k^2\}$$

即 $Y_1,Y_2,\cdots,Y_n$ 相互独立地服从 N(0,1) 分布.

习 题 5

1. (1) $e^{-5\lambda}(5\lambda)^4/4!$.
(2) $\{e^{-5\lambda}(5\lambda)^4/4!\}\cdot\{e^{-2.5\lambda}(2.5\lambda)^2/2!\}\cdot\{e^{-4.5\lambda}(4.5\lambda)^3/3!\}$
(3) $e^{-7\lambda}(7\lambda)^5/5!$ (4) 5λ (5) 5λ (6) $4+5\lambda$

2. 令最初的公共汽车出发时刻为时间的原点，并令时刻 t 为止所到乘客数为 $N(t)$，又令下一辆公共汽车的到达时刻为 T. 则 $E(N(t))=t$, $E(N^2(t)=t+t^2$. $E(N(T))=E(T)=10$, $\mathrm{Var}(N(T))=E(N^2(t))-\{E(N(T)\}^2$, $E(N^2(6))=\frac{1}{4}\int_8^{12}E(N^2(t))\mathrm{d}t=\frac{1}{4}[\frac{t^2}{2}+\frac{t^3}{3}]_8^{12}=10+\frac{304}{3}$. 所以 $\mathrm{Var}(N(t))\approx 11.3$

3. 因为 $N(t+s)-N(t)$ 与 $N(t)$ 独立，$E[\{N(t+s)-N(t)\}N(t)]=E[N(t+s)-N(t)]\cdot E[N(t)]=\lambda s\cdot\lambda t$ 又由于 $E[\{N(t+s)-N(t)\}N(t)]=E[N(t+s)N(t)]-E[N^2(t)]$. 所以

$$\begin{aligned}\mathrm{Cov}(N(t-s),N(t))=&E[N(t+s)N(t)]-E[N(t+s)]E[N(t)]\\=&\lambda s\cdot\lambda t+\{\lambda t+(\lambda t)^2\}-\lambda(t+s)\lambda t=\lambda t\end{aligned}$$

4. 由表 3-1(b), $E(Z)=3/5$, $\mathrm{Var}(Z)=1/25$, 故 $E(Y)=40(3/5)+10=34$, $\mathrm{Var}(Y)=40^2(1/25)=64$. 又由于 1 小时内到来的车辆数的数学期望为 30 辆，故由式 (5.22) 及式 (5.23), 加油总量的数学期望及方差分别为 $30\cdot 34=1020[l]$, $30\cdot(64+34^2)=36600[l^2]$.

5. 由 $p_n(S+h)=p_n(s)(1-\lambda h)+p_{n-1}(s)\lambda h+o(h)$

$$\frac{p_n(s+h)-p_n(s)}{h}=-\lambda p_n(s)+\lambda p_{n-1}(s)+\frac{o(h)}{h}$$

令 $h \to 0$, 则

$$p_n'(s) = -\lambda p_n(s) + \lambda p_{n-1}(s) \quad (n \geq 1) \tag{5.27}$$

特别地, $n = 0$ 时候,

$$p_0'(s) = -\lambda p_0(s)$$

其解为 $p_0(s) = p_0(0)e^{-\lambda s} = e^{-\lambda s}$. 在式 (5.27) 中令 $n = 1$, 我们可以求出 $p_1(s) = \lambda s e^{-\lambda s}$. 同理, 我们还可以求出

$$p_n(s) = \frac{(\lambda s)^n}{n!}e^{-\lambda s}$$

6. (1) 略. (2) $p_0(t+h) = p_0(t)(1-\lambda h) + p_1(t)\mu h(1-\lambda h) + o(h)$.

(3)

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}p_n(t) = -(\mu+\lambda)p_n(t) + \mu p_{n+1}(t) + \lambda p_{n-1}(t) & (n \geq 1) \\ \dfrac{\mathrm{d}}{\mathrm{d}t}p_0(t) = -\lambda p_0(t) + \mu p_1(t) \end{cases}$$

(4)

$$\begin{cases} -(\mu+\lambda)p_n + \mu p_{n+1} + \lambda p_{n-1} = 0 \quad (n \geq 1) & (*) \\ -\lambda p_0 + \mu p_1 = 0 & (**) \end{cases}$$

(5) 式 (*) 可写成 $-\lambda p_n + \mu p_{n+1} = -\lambda p_{n-1} + \mu p_n$ 的形式. 再由式 (**), 我们知道对于所有 n, $-\lambda p_n + \mu p_{n+1} = 0$. 即 $p_{n+1} = (\lambda/\mu)p_n$, $\{p_n\}$ 是公比为 $\rho = \lambda/\mu$ 的等比数列, 且 p_n 可表示成 $p_n = \rho^n p_0$ 的形式. 为了使 $\sum\limits_{n=0}^{\infty} p_n = 1$, 公比 ρ 应该小于 1. 此时 $\sum_{n=0}^{\infty} p_n = p_0/(1-\rho)$, $p_0 = 1-\rho$, $p_n = (1-\rho)\rho^n$.

习题 6

1. $\widetilde{f}(z) = (\frac{2}{z+2})^3$, $\widetilde{m}(z) = \frac{8}{z^2(z^2+6z+12)} = \frac{1}{3}[\frac{2}{z^2} - \frac{1}{z} +$

$\frac{(z+3)+1}{(z+3)^2+3}]$.

$$m(t)=\frac{1}{3}[2t-1+e^{-3t}(\cos\sqrt{3}t+\frac{1}{\sqrt{3}}\sin\sqrt{3}t)]$$
$$=\frac{1}{3}[2t-1+\frac{2}{\sqrt{3}}e^{-3t}\cos(\sqrt{3}t-\frac{\pi}{6})]$$

2. 因为 $m(t)$ 的拉普拉斯变换 $\widetilde{m}(z)=\lambda/z^2$, 故由式 (6.13), $\widetilde{f}(z)=z\widetilde{m}(z)/(1+z\widetilde{m}(z))=\lambda/(z+\lambda)$. 最后通过逆变换得到 $f(t)=\lambda e^{-\lambda t}$

3. 通话结束时刻的序列为再生过程. 再生时间间隔的均值 $u=1/\lambda+E(Y_1)$, 故由式 (6.19), 即可得到所需结果.

4. (1) 令最初电话打来时刻为 T_1, 则

$$m(t)=\int_0^t E[N(t)|T_1=s]\lambda e^{-\lambda s}\mathrm{d}s$$

又由于此电话通话至时刻 t 的概率为 $1-G(t-s)$. 故 $E[N(t)|T_1=s]=1\times[1-G(t-s)]+m(t-s)$

(2) $\widetilde{m}(z)=[\frac{1}{z}-\widetilde{G}(z)]\frac{\lambda}{z+\lambda}+\widetilde{m}(z)\frac{\lambda}{z+\lambda}\Rightarrow$
$\widetilde{m}(z)=\lambda[\frac{1}{z}-\widetilde{G}(z)]\frac{1}{z}\Rightarrow m(t)=\lambda\int_0^t[1-G(t)]\mathrm{d}t$

5. (1), (2) 略. (3) $\widetilde{m}(z)=\frac{1}{1-a\widetilde{f}(z)}[\frac{1-\widetilde{f}(z)}{z}]$

(4) $z\to 0$ 时, $\frac{1-\widetilde{f}(z)}{z}$ 接近 0/0, 于是根据洛毕达法则, 其极限值为
$-\frac{\mathrm{d}}{\mathrm{d}z}\widetilde{f}(z)|_{z=0}=\mu$. 最后我们得到 $L=\mu/(1-a)$. 6. (1) 令替换间隔为 D, 则 $E(D)=\int_0^\infty P\{D>u\}\mathrm{d}u=$
$\int_0^T[1-F(u)\mathrm{d}u]$. 再考虑利用此式与 (6.20) 式.

(2) 略.

(3) 因故障替换的间隔的数学期望为

$$\int_0^{\infty}[1-G(t)]\mathrm{d}t=\sum_{k=0}^{\infty}\int_{k\tau}^{(k+1)\tau}[1-G(t)]\mathrm{d}t$$

$$=\sum_{k=0}^{\infty}[1-F(\tau)]^k\int_0^{\tau}[1-F(u)]\mathrm{d}u$$

$$=\int_0^{\tau}[1-F(u)]\mathrm{d}u/F(\tau)$$

再考虑利用上式与式 (6.20) 即可.

(4) 令 $F(u)=1-e^{-\lambda u}$, 则 $C(\tau)=\lambda[c_1(1-e^{-\lambda\tau})+c_2e^{-\lambda\tau}]/(1-e^{-\lambda\tau})$. 即它为 τ 的单调递减函数.

7. (1) 令最初的再生时刻为 T_1, 则

$$m_2(t)=\int_0^{\infty}E[N^2(t)|T_1=s]\mathrm{d}F(s)=\int_0^{t}E[\{1+N(t-s)\}^2]\mathrm{d}F(s)$$

$$=F(t)+2\int_0^{t}m(t-s)\mathrm{d}F(s)+\int_0^{t}m_2(t-s)\mathrm{d}F(s)$$

对它进行拉普拉斯变换, 则 $\tilde{m}_2(z)=\tilde{f}(z)/z+2\tilde{m}(z)\tilde{f}(z)+\tilde{m}_2(z)\tilde{f}(z)$. 再根据式 (6.13), $\tilde{m}_2(z)=\tilde{m}(z)+2\tilde{m}(z)\cdot z\tilde{m}(z)$. 最后通过逆变换即可得到所需的结果.

(2) 根据第 3 节推出的展式, 我们可以得到 $\tilde{m}_2(z)$ 的展式, 最后再通过逆变换可以求出 $m_2(t)$.

(3) $N(t)\geq n\Leftrightarrow T_n\leq t$, 且 T_n 为相互独立地服从同一分布的 n 个随机变量之和, 故可用中心极限定理. 要注意 $E(N(t))=t/\mu+O(1)$, $\mathrm{Var}(N(t))=m_2(t)-\{m(t)\}^2=t\sigma^2/\mu^3+O(1)$

8. (1),(2) 略.

(3) 只需证出 $F^*(z)-zF^*(z)=f^*(z)$.

(4) 利用式 (6.36) 和 (6.37).

(5) $f^*(z)=pz/(1-qz)$, $m^*(z)=pz/(1-z)^2=pz(1+z+z^2+\cdots)^2=pz(1+2z+3z^2+\cdots)$, $m(k)=kp$.

习 题 7

1. (a) $\{1,3,4\}$: 正常返的, $d=1$; $\{2\}$: 非常返的, $d=1$; $\{5\}$: 非常返的, $d=1$. (b) $\{1,2,5\}$: 正常返的, $d=1$; $\{3,4\}$: 正常返的, $d=1$. (c) $\{1,5,6,7\}$: 非常返的, $d=2$; $\{2,3\}$: 正常返的, $d=2$; $\{4\}$: 非常返的, $d=1$. (d) $\{1\}$: 非常返的, $d=1$; $\{2,5,6\}$: 正常返的, $d=2$; $\{3\}$: 正常返的, $d=1$; $\{4,7\}$: 正常返的, $d=2$.

2. (1) $T=\{2,5\}$, $S-T=\{1,3,4\}$. (2) 略.

(3)

$$\begin{array}{c} \\ 2 \\ 5 \end{array} \begin{array}{c} \begin{array}{cc} 2 & 5 \end{array} \\ \begin{bmatrix} 0.3 & 0 \\ 0.5 & 0.4 \end{bmatrix} \end{array} = Q$$

$$N=(I-Q)^{-1}=\frac{1}{0.42}\begin{bmatrix} 0.6 & 0 \\ 0.5 & 0.7 \end{bmatrix}$$

(4)

$$\begin{bmatrix} m(2) \\ m(5) \end{bmatrix} = N\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{0.42}\begin{bmatrix} 0.6 \\ 1.2 \end{bmatrix} = \begin{bmatrix} 10/7 \\ 20/7 \end{bmatrix}$$

3. (1) 如果 $l=0$ 或为 K, 则

$$f(i,l)=pf(i+1,l)+qf(i-1,l) \tag{i}$$

这里,

$$f(0,0)=f(K,K)=1, f(0,K)=f(K,0)=0 \tag{ii}$$

差分方程式 (i) 的特征方程为 $\lambda=p\lambda^2+q$, 其解为 $\lambda=1$, $\lambda=q/p$. 所以, 利用边界条件 (ii) 可以求出 A, B. 于是

$$f(i,0)=\frac{(q/p)^i-(q/p)^K}{1-(q/p)^K}, \qquad f(i,K)=\frac{1-(q/p)^i}{1-(q/p)^K}$$

(2) 在边界条件 $m(0)=m(K)=0$ 下, 解出 $m(i)=1+pm(i+1)+qm(i-1)$ 即可. 答案为 $m(i)=\dfrac{K}{p-q}\cdot\dfrac{1-(q/p)^i}{1-(q/p)^K}-\dfrac{i}{p-q}$.

4. (1) 假定 $i \in T, j \in S-T$, 则 $f(i,i) = P(i,j) + \sum_{k\in T} P(i,k)f(k,j)$. 如用矩阵表示，则 $\boldsymbol{F} = \boldsymbol{R} + \boldsymbol{QF}$, 即 $\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}\boldsymbol{R} = \boldsymbol{NR}$.

(2) $m^{(2)}(i) = \sum_{j\in S-T} P(i,j) \times 1^2 + \sum_{k\in T} P(i,k)E[(1+A_k)^2]$. 这里，$A_k$ 为从 k 出发到达 $S-T$ 所需的转移次数，$E[(1+A_k)^2] = 1 + 2m(k) + m^{(2)}(k)$. 因而 $m^{(2)}(i) = 1 + 2\sum_{k\in T} P(i,k)m(k) + \sum_{k\in T} P(i.k)m^{(2)}(k)$. 用矩阵表示，则为 $\boldsymbol{m}^{(2)} = \boldsymbol{1} + 2\boldsymbol{Qm} + \boldsymbol{Qm}^{(2)}$. 再由式 (7.40) $\boldsymbol{m} = \boldsymbol{N1}$, 我们即可得到所需结果.

(3)

$$\boldsymbol{F} = \boldsymbol{NR} = \frac{1}{1-pqr}\begin{bmatrix} pqr & r & rq \\ pq & pqr & q \\ p & pr & pqr \end{bmatrix}, \quad \boldsymbol{m}^{(2)} = \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix}$$

5. (1) 略. (2) $P^{(2m)}(j,j) \geq \{P^{(m)}(j,j)\}^2 > 0$.

(3) $P^{(l+m+n)}(i,i) \geq P^{(l)}(i,j)P^{(m)}(j,j)P^{(n)}(j,i) > 0$. 其他类同.

(4) 因为 $(l+m+n)$, $(l+2m+n)$ 均能被 $d(i)$ 整除，故其差也能被 $d(i)$ 整除.

(5) 上式对于任意满足 $P^{(m)}(j.j) > 0$ 的 m 均成立，故 $d(j)$ 能被 $d(i)$ 整除.

(6) i 和 j 互换，则 $d(i)$ 亦能被 $d(j)$ 整除. 所以 $d(i) = d(j)$.

6. 因为具有两个等价类，所以不存在极限分布. 对于每个等价类，首先求其平稳分布

$$\left.\begin{array}{l} \pi(1) = 0.5\pi(1) + 0.5\pi(5) \\ \pi(2) = 0.3\pi(1) + 0.7\pi(2) \\ \pi(5) = \pi(2) \\ \pi(1) + \pi(2) + \pi(5) = 1 \end{array}\right\} \Rightarrow \pi(1) = \pi(2) = \pi(5) = 1/3$$

$$\left.\begin{array}{l} \pi(3) = 0.4\pi(3) + 0.6\pi(4) \\ \pi(4) = \pi(3) \\ \pi(3) = \pi(4) = 1 \end{array}\right\} \Rightarrow \pi(3) = \pi(4) = 1/2$$

对于 $0 \le \omega \le 1$ 的任意 ω, $\pi(\omega) = (\omega/3, \omega/3, (1-\omega)/2, (1-\omega)/2, \omega/3)$ 为平稳分布．所以，当依概率 ω 从等价类 $\{1,2,5\}$ 或依概率 $(1-\omega)$ 从等价类 $\{3,4\}$ 出发时． $\lim_{n\to\infty} P(X_n = i) = \pi_\omega(i)$.

7. (1) 略． (2) 因为这个系统是既约的，且周期为 2, 故不存在极限分布，但存在平稳分布．其平衡方程式为

$$\pi(0) = (1/a)\pi(1), \qquad \pi(a) = (1/a)\pi(a-1),$$
$$\pi(j) = (1 - \frac{j-1}{a})\pi(j-1) + \frac{j+1}{a}\pi(j+1) \quad (j = 1, 2, \cdots, a-1)$$

解之， $\pi(j) = \binom{a}{j} 2^{-a}$, 即平稳分布服从均值为 $a/2$ 的二项分布．

8. 此系统是正常返的，故存在极限分布，且它是平衡方程式的唯一解．很显然，其解为 $\pi = (1/a, 1/a, \cdots, 1/a)$.

索 引

六画

七画

八画

九画

十画

十一画

十二画

十三画

十四画

十五画